Julia, Gaston
Introduction mathematique
aux theories Quantiques, 1936

OUVRAGES DU MÊME AUTEUR

Leçons sur les Fonctions uniformes à point singulier essentiel isolé, *professées au Collège de France*, rédigées par P. Flamant, Agrégé préparateur à l'École Normale supérieure. Un volume in-8 raisin (25-16) de 152 pages.. **25** fr.

Éléments de Géométrie infinitésimale. (*Cours de la Faculté des Sciences de Paris.*) Volume in-8 raisin de 242 p., avec 15 fig... **45** fr.

Exercices d'Analyse.

Tome I. In-8 (25-16) de vii-454 pages, avec 106 figures...... **80** fr.
Tome II. In-8 (25-16) de vi-344 pages, avec 86 figures....... **70** fr.
Tome III. In-8 (25-16) de iv-287 pages, avec 37 figures....... **60** fr.
Tome IV. In-8 (25-16) de 230 pages...................... **60** fr.

Principes géométriques d'Analyse (première Partie), recueillis et rédigés par MM. Brelot et René de Possel, Agrégés de l'Université, anciens Élèves de l'École Normale supérieure. (*Cahiers scientifiques*, fascicule VI.) Un volume in-8 raisin (25-16) de 116 pages, avec 35 figures..... **25** fr.

Principes géométriques d'Analyse (deuxième Partie), recueillis et rédigés par André Magnier, Élève de l'École Normale supérieure. (*Cahiers scientifiques*, fascicule XI.) Un volume in-8 raisin (25-16) de vii-120 pages, avec 37 figures.. **40 fr.**

Leçons sur la représentation conforme des aires simplement connexes. (*Cahiers scientifiques*, fascicule VIII.) Un volume in-8 raisin (25-16) de 112 pages, avec 39 figures....................... **30** fr.

Leçons sur la représentation conforme des Aires multiplement connexes. Leçons recueillies et rédigées par MM. Georges Bourion et Jean Leray, Docteurs ès sciences. (*Cahiers scientifiques*, fascicule XIV.) In-8 (25-16) de vi-96 pages avec 36 figures................... **28** fr.

Essai sur le développement de la Théorie des fonctions de variables complexes. (*Conférence faite au Congrès international de Mathématiques à Zurich* (1932) *et Extrait des Comptes rendus du Congrès.*) In-8 (23-14), viii-54 pages....................... **12** fr.

Cours de Cinématique, rédigé par Jean Dieudonné, Élève de l'École Normale supérieure. (*Cours de la Faculté des Sciences de Paris.*) Un volume in-8 carré (23-14) de 150 pages, avec 52 figures........ **25** fr.

Exercices de Mécanique, par MM. H. Béghin et G. Julia, in-8 raisin (25-16)

Tome I. Fascicule I. Volume de viii-328 pages, avec 141 figures.... **80** fr.
Tome I. Fascicule II. Volume de 240 pages, avec 86 figures........ **60** fr.

CAHIERS SCIENTIFIQUES

PUBLIÉS SOUS LA DIRECTION DE M. GASTON JULIA

FASCICULE XVI

INTRODUCTION MATHÉMATIQUE

AUX

THÉORIES QUANTIQUES

PAR

Gaston JULIA

MEMBRE DE L'INSTITUT

PROFESSEUR A LA FACULTÉ DES SCIENCES DE PARIS

Leçons rédigées par J. DUFRESNOY

Élève à l'École Normale supérieure

PREMIÈRE PARTIE



PARIS

GAUTHIER-VILLARS, ÉDITEUR

LIBRAIRE DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE

55, Quai des Grands-Augustins, 55

1936

PRÉFACE.

Le présent fascicule contient la première partie de leçons données à la Sorbonne au printemps de 1935. Il doit être suivi d'un ou de plusieurs autres fascicules, dont l'ensemble formera une *introduction mathématique aux théories quantiques*.

Le présent fascicule expose, à partir des éléments, les propriétés des espaces vectoriels à n dimensions et de leurs transformations linéaires. Géométriquement, on obtient une théorie des *opérateurs linéaires*, qui se traduit analytiquement par une théorie des *matrices*. Le choix des coordonnées aptes à fournir la matrice la plus simple d'un opérateur déterminé conduit naturellement à la *réduction des matrices*, des formes qui leur sont attachées, et à l'étude du *spectre*. En géométrie affine on obtient ainsi les réduites de *Jordan;* en géométrie métrique ou unitaire on obtient les réduites de *Schur*, et, lorsque la matrice est *hermitienne*, on a la réduite diagonale que Lagrange et Gauss avaient donnée pour les formes quadratiques.

La matière de l'*Introduction* que nous préparons, se trouve déjà exposée, plus ou moins complètement, à des points de vue et sous des formes assez divers, dans plusieurs bons ouvrages, parmi lesquels nous citerons surtout :

F. Riesz, *Les systèmes d'équations linéaires à une infinité d'inconnues* (1913);

H. Weyl, *Gruppentheorie und Quantenmechanik* (1928);

A. Wintner, *Spektraltheorie der unendlichen Matrizen* (1929);

M. H. Stone, *Linear transformations in Hilbert Space and their applications to Analysis* (1932) ;

J. VON NEUMANN, *Mathematische Grundlagen der Quantenmechanik* (1932).

La lecture des livres précédents nous a été fort utile en période d'élaboration de ces leçons ; utiles aussi le bel article de l'Encyclopédie allemande que l'on doit à MM. HELLINGER et TOEPLITZ, *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, et l'excellent Traité de MM. COURANT et HILBERT, *Methoden der mathematischen Physik*.

Il nous paraît cependant qu'à garder simultanément présentes à l'esprit la notion géométrique d'opérateur et les représentations diverses de cet opérateur, par des matrices correspondant aux divers systèmes de coordonnées offerts par la géométrie de l'espace étudié, l'exposition gagne en clarté et en simplicité. C'est dans ce sens qu'on a essayé de mener ici l'exposition et qu'on espère, par fusion du géométrique et de l'analytique, avoir réussi à éclaircir ou simplifier plus d'une démonstration et à faciliter l'acquisition des notions qui font l'objet de ces leçons.

Ces leçons ont été faites à la demande de quelques physiciens, devant des physiciens et des mathématiciens. Pour les mathématiciens qui voudraient approfondir l'étude des questions envisagées ici, il est bien évident que notre livre n'est qu'une introduction à poursuivre l'étude des mémoires originaux, en tête desquels figurent les beaux mémoires de M. Hilbert, sur les équations intégrales.

Il m'est agréable de remercier ici M. Dufresnoy, élève de l'École Normale Supérieure, qui a bien voulu rédiger ces leçons, et M. Winter qui, après avoir assisté à ces leçons, a bien voulu revoir et critiquer la rédaction « en physicien ». La maison Gauthier-Villars n'a rien négligé pour assurer à ce livre la meilleure présentation; qu'elle soit, elle aussi, chaleureusement remerciée.

GASTON JULIA.

INTRODUCTION MATHÉMATIQUE

AUX

THÉORIES QUANTIQUES

APERÇU HISTORIQUE.

I. — GÉNÉRALITÉS.

1. L'origine des théories mathématiques se trouve dans l'étude de problèmes physiques. Il est, à première vue, surprenant qu'une théorie ainsi ébauchée, puis développée indépendamment de toute considération extérieure aux mathématiques, vienne par la suite rendre au physicien les plus grands services. C'est pourtant de cette façon que se construisent parallèlement la physique et les mathématiques.

2. Les problèmes relatifs aux vibrations des systèmes conduisent à des équations différentielles et à des équations aux dérivées partielles; comme autre exemple de problème physique conduisant au même genre de questions mathématiques, on doit citer le potentiel électrique ou magnétique. La détermination d'un équilibre électrique, par exemple, nous conduit à rechercher les solutions de l'équation

$$\Delta U = 0,$$

U désignant une fonction d'un point de l'espace (fonction qui représente le potentiel électrostatique), Δ désignant le laplacien

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

solutions qui doivent prendre des valeurs données sur des surfaces données. C'est le célèbre problème de Dirichlet. Plus généralement on sera conduit à la même question avec l'équation

$$\Delta U = f,$$

f étant une fonction donnée $\left(\text{par exemple, dans le problème précédent, } \frac{f}{4\pi} \text{ pourra représenter une densité de charges électriques, imposées à l'avance}\right)$.

Pour le problème des cordes vibrantes, étudié depuis Bernoulli, d'Alembert etc., si l'on cherche à représenter l'élongation z de la corde tendue, par une fonction de la forme $z = u(s).\, v(t)$, s étant l'abscisse et t le temps, z satisfaisant à l'équation

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial s^2}$$

(qui exprime que l'accélération de l'élément de corde d'abscisse s est due aux tensions provenant des éléments voisins), on obtient

$$c^2 \frac{d^2 u}{ds^2} v = u \frac{d^2 v}{dt^2},$$

soit

$$\frac{1}{c^2} \frac{d^2 v(t)}{dt^2} \frac{1}{v(t)} = \frac{d^2 u(s)}{ds^2} \frac{1}{u(s)} = -\lambda,$$

λ étant une constante indépendante de s et de t, ce qui conduit à l'équation

$$\frac{d^2 u}{ds^2} + \lambda u = 0,$$

dont on cherche les solutions $u(s)$ s'annulant aux extrémités de la corde. C'est ce problème des cordes vibrantes qui a conduit à la considération des séries de Fourier et à la théorie des systèmes orthogonaux de fonctions.

Dans l'étude des vibrations de membranes, on recherchera de même des solutions $u(x, y)$ de l'équation

$$\Delta u + \lambda u = 0,$$

qui soient nulles sur le contour sans être identiquement nulles.

Ces problèmes ne seront résolubles que pour certaines valeurs de la constante λ, dépendant des extrémités de la corde ou du contour de la membrane. C'est évident pour la corde : soit $\lambda > 0$, on peut poser $\lambda = \omega^2$ et l'on obtient la solution

$$u = A \cos \omega s + B \sin \omega s.$$

Annulons pour $s = 0$ et $s = l$; il vient $A = 0$, et si l'on veut avoir $B \neq 0$

il faut $\sin \omega l = 0$, soit $\omega = \frac{k\pi}{l}$, $\lambda = k^2 \frac{\pi^2}{l^2}$, k étant un entier arbitraire. Si $\lambda < 0$ on aurait à considérer les cosinus et sinus hyperboliques; on verrait sans peine que le problème ne serait plus résoluble.

Ces valeurs de λ sont dites *valeurs propres* ou bien valeurs *fondamentales*, ou encore valeurs *caractéristiques*. Les solutions correspondant à de telles valeurs seront dites, de même, propres, fondamentales ou caractéristiques.

Ce problème joue un rôle fondamental en acoustique et en spectroscopie. La constante λ reçoit une interprétation physique très importante : elle est proportionnelle à l'énergie du système en vibration, qui apparaît ainsi comme susceptible de prendre uniquement des valeurs discrètes.

3. Les problèmes dont il vient d'être question sont des *problèmes linéaires*; cela veut dire que si l'on en a deux solutions, leur somme est une solution et que si l'on en a une solution, celle-ci multipliée par une constante quelconque est encore une solution.

En physique on rencontre fréquemment de tels problèmes linéaires; cela tient moins à un fait physique qu'à la relative facilité de la résolution de ces problèmes : ce sont presque les seuls que l'on ait étudiés.

4. La démonstration rigoureuse de l'existence, et le cas échéant de l'unicité, de la solution de tels problèmes, ne s'appuyant pas sur des considérations physiques, est difficile et récente. Après Riemann, Neumann, Schwarz, Poincaré ont fait des travaux sur le problème de Dirichlet; pour les cordes vibrantes, la première solution satisfaisante date de Fourier, puis de Dirichlet; quant aux autres problèmes, la démonstration de l'existence de valeurs caractéristiques λ a été partiellement faite par Schwarz (qui a mis en évidence la solution la plus petite), par M. Picard (qui a montré l'existence de la valeur caractéristique suivante), puis a été complètement réalisée par Poincaré (1). Fredholm a retrouvé tous ces résultats par une méthode différente.

Considérons l'équation

$$\Delta U + \lambda U = 0;$$

on peut lui donner une autre forme, qui sera d'ailleurs entièrement équivalente lorsque le Δ de la solution existera. Considérons une distribution de matière, de densité μ, dans une certaine aire; elle crée un

(1) *Voir* Poincaré, *Rendiconti del Circolo Matematico di Palermo*, 1894.

potentiel

$$U(M) = \int\int \mu(P) \log \frac{1}{MP}\, d\sigma_P.$$

Hors des masses $\Delta U = 0$, tandis que dans les masses $\Delta U = -2\pi\mu$ d'après la formule de Poisson. Rappelons, en passant, que l'on peut faire les mêmes considérations dans l'espace, mais on devra alors remplacer $\log\frac{1}{MP}$ par $\frac{1}{MP}$ et le coefficient de la formule de Poisson sera 4π au lieu de 2π.

Soit alors à résoudre

$$\Delta U + \lambda U = f.$$

Si l'on pose

$$U(M) = \int\int \mu(P) \log \frac{1}{MP}\, d\sigma_P,$$

il vient

$$\Delta U(M) = -2\pi\mu(M) = f(M) - \lambda U(M).$$

D'où l'équation

$$U(M) = -\frac{1}{2\pi}\int\int [f(P) - \lambda U(P)] \log \frac{1}{MP}\, d\sigma_P,$$

dont on est ramené à chercher une solution U; avec les mêmes conditions de contour, ces deux problèmes sont équivalents; cette équation est une équation intégrale qui peut s'écrire

$$U(M) - \frac{\lambda}{2\pi}\int\int K(M, P)\, U(P)\, d\sigma_P = F(M),$$

dans laquelle U est la fonction inconnue. Remarquons que dans le cas particulier où $f = 0$, on a

$$F = -\frac{\lambda}{2\pi}\int\int f(P) \log \frac{1}{MP}\, d\sigma_P = 0,$$

et nous retrouvons encore une équation linéaire. Si nous avons un second membre F, il est facile de s'assurer que la solution générale d'une telle équation est la somme d'une solution particulière et de la solution générale de l'équation sans second membre.

Avec une seule variable, nous serions de même conduit à une équation du type

$$\varphi(s) + \lambda \int_a^b K(s, t)\, \varphi(t)\, dt = f(s),$$

dans laquelle φ est la fonction inconnue. La constante λ s'introduira facilement si l'on part des cordes vibrantes; par contre, si l'on part du

problème de Dirichlet $\Delta U = 0$, et qu'on essaie de le résoudre par un potentiel de double couche de densité inconnue μ, tel que U, lorsqu'on s'approche du contour, tende vers les valeurs imposées au contour, on est conduit pour déterminer μ à une équation du même type, mais ne contenant pas de constante λ.

5. Fredholm a fait l'étude des équations

$$\varphi(s) + \lambda \int_a^b K(s, t)\varphi(t)\,dt = f(s),$$

lorsque K (*noyau*) est une fonction continue de ses deux arguments et même dans des cas un peu plus généraux ([1]).

Avant lui, Le Roux et Volterra avaient résolu la question dans le cas particulier où $K(s, t)$ est nul lorsque $t > s$; l'équation intégrale s'écrit alors

$$\varphi(s) + \lambda \int_a^s K(s, t)\varphi(t)\,dt = f(s),$$

et la méthode des approximations successives s'applique pour toute valeur de λ :

$$\varphi_0(s) = f(s),$$

$$\varphi_1(s) = f(s) - \lambda \int_a^s K(s, t)\varphi_0(t)\,dt,$$

. . . . ,

$$\varphi_n(s) = f(s) - \lambda \int_a^s K(s, t)\varphi_{n-1}(t)\,dt;$$

. . . .

φ_n tend vers φ, solution de l'équation. Tandis que dans le cas envisagé par Fredholm la méthode ne réussit que si λ est suffisamment petit : d'une façon plus précise, si $|\lambda|$ est inférieur au module de la racine de plus petit module de $D(\lambda) = 0$, équation que nous allons définir.

Revenons aux travaux de Fredholm. Ils montrent qu'une valeur de λ permet en général la résolution; la solution se présente sous la forme

$$\varphi(s) = f(s) - \lambda \int_a^b \mathcal{K}(s, t; \lambda) f(t)\,dt.$$

$\mathcal{K}(s, t; \lambda)$ étant une fonction que l'on sait calculer; on l'appelle *noyau résolvant de* K, et elle se présente sous forme d'une fonction méro-

(1) *Voir* FREDHOLM, *Acta mathematica*, 1901.

morphe en λ ou, d'une façon plus précise, comme quotient de deux fonctions entières

$$\mathcal{K}(s, t; \lambda) = \frac{D(s, t; \lambda)}{D(\lambda)}.$$

La fonction $D(\lambda)$, que l'on appelle le *déterminant* de l'équation, se forme uniquement en connaissant $K(s, t)$; c'est à rapprocher de la résolution d'un système d'équations linéaires : on sait que le déterminant du système ne dépend que des premiers membres. Plusieurs cas sont possibles :

1° Si λ n'annule pas $D(\lambda)$, une solution et une seule (rapprocher du système de Cramer d'équations linéaires).

2° Sinon, que fait-on dans le cas du système d'équations

$$\sum_k a_{ik} x_k = y_i ?$$

On considère les premiers membres; on cherche les formes linéairement indépendantes et les relations linéaires existant avec les autres formes, c'est-à-dire les systèmes de nombres λ_i, non tous nuls, satisfaisant à

$$(1) \qquad \sum_i \lambda_i a_{ik} = 0,$$

pour tous k, et, par application du théorème de Rouché, on en déduit que *la condition nécessaire et suffisante pour que le système présente des solutions est qu'il existe entre les seconds membres toute combinaison linéaire qui existe entre les premiers*, soit

$$(2) \qquad \sum_i \lambda_i y_i = 0,$$

pour tous les systèmes λ_i de valeurs précédemment trouvés.

Ici, au *système transposé* (1) correspond *l'équation homogène associée*

$$\varphi(s) + \lambda \int_a^b K(t, s)\, \varphi(t)\, dt = 0.$$

De même que le système (1) présentait toujours un nombre fini de solutions linéairement indépendantes, cette équation présente toujours un nombre fini de solutions

$$\varphi_1, \quad \varphi_2, \quad \ldots, \quad \varphi_k,$$

linéairement indépendantes.

De même que l'équation (2) exprimait que le vecteur Y de composantes $y_1, y_2, \ldots, y_n$ était perpendiculaire à tous les vecteurs Λ de composantes $\lambda_1, \lambda_2, \ldots, \lambda_n$, c'est-à-dire, d'une façon abrégée, que le second membre doit être orthogonal aux solutions du système transposé, on trouve ici, comme condition nécessaire et suffisante pour que l'équation intégrale ait des solutions, que f soit orthogonal à φ_1, $\varphi_2, \ldots, \varphi_k$, c'est-à-dire

$$\int_a^b \varphi_i(t) f(t)\, dt = 0 \qquad \text{pour} \quad i = 1, 2, \ldots, k.$$

II. — PREMIÈRE MÉTHODE DE HILBERT.

1. Quelque temps après Fredholm, Hilbert a repris la question à un double point de vue algébrique. Considérons sa première méthode. Elle nous montre le parallélisme du problème avec celui des équations linéaires.

Nous nous proposons de résoudre

$$\varphi(s) + \lambda \int_a^b K(s, t)\, \varphi(t)\, dt = f(s).$$

Le premier membre de cette relation est une *fonctionnelle* de la fonction φ, que nous désignerons par $\mathfrak{G}(\varphi)$, c'est-à-dire une expression que nous saurons calculer si nous connaissons la *fonction* φ; c'est donc là une généralisation de la notion de fonction. Cette *fonctionnelle* est *linéaire*, c'est-à-dire que l'on a

$$\mathfrak{G}(a\varphi) = a\mathfrak{G}(\varphi),$$

si a désigne une constante et

$$\mathfrak{G}(\varphi_1 + \varphi_2) = \mathfrak{G}(\varphi_1) + \mathfrak{G}(\varphi_2).$$

Nous allons évaluer le premier membre d'une façon approchée en remplaçant l'intégrale par une somme; à cet effet nous diviserons l'intervalle ab, en intervalles partiels de longueur $\delta = \frac{b-a}{n}$, par les points $x_1, x_2, \ldots, x_n$. Nous aurons alors la relation *approchée*

$$\varphi(s) + \lambda \sum_i K(s, t_i)\, \varphi(t_i)\, \delta = f(s) \qquad (t_i = x_i),$$

relation qui sera en particulier vérifiée lorsque nous donnerons à s les

valeurs $x_1, x_2, \ldots, x_n$; d'où les relations

$$\varphi(s_j) + \lambda \sum_i K(s_j, t_i)\varphi(t_i)\delta = f(s_j) \qquad \text{pour } j = 1, 2, \ldots, n \qquad (s_j = x_j).$$

Si nous posons

$$\begin{aligned} \varphi(s_j) &= \varphi_j, \\ \delta K(s_j, t_i) &= K_{ji}, \\ f(s_j) &= f_j, \end{aligned}$$

ces relations prendront la forme

$$\varphi_j + \lambda \sum_i K_{ji}\varphi_i = f_j,$$

soit, avec les notations développées,

$$\begin{array}{l} (1 + \lambda K_{11})\varphi_1 + \lambda K_{12}\ \varphi_2 + \ldots + \lambda K_{1n}\ \varphi_n = f_1, \\ \lambda K_{21}\ \varphi_1 + (1 + \lambda K_{22})\varphi_2 + \ldots + \lambda K_{2n}\ \varphi_n = f_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ \lambda K_{n1}\ \varphi_1 + \lambda K_{n2}\ \varphi_2 + \ldots + (1 + \lambda K_{nn})\varphi_n = f_n. \end{array}$$

Les φ_i sont ainsi déterminés par un système d'équations linéaires; nous avons donc calculé la valeur approchée de la fonction φ en n points intermédiaires de l'intervalle ab.

Cette idée très simple, employée par Hilbert, s'était montrée féconde pour d'autres problèmes importants; elle est, en effet, à rapprocher de la méthode de Cauchy-Lipschitz pour les équations différentielles

$$\frac{dy}{dx} = f(x, y),$$

ou, plus simplement encore, de la recherche des primitives d'une fonction en appliquant la méthode précédente à

$$\frac{dy}{dx} = f(x).$$

Pour pouvoir résoudre le système linéaire d'une façon unique, il faut que

$$\Delta(\lambda) = \begin{vmatrix} 1 + \lambda K_{11} & \lambda K_{12} & \ldots & \lambda K_{1n} \\ \lambda K_{21} & 1 + \lambda K_{22} & \ldots & \lambda K_{2n} \\ \ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots \\ \lambda K_{n1} & \lambda K_{n2} & \ldots & 1 + \lambda K_{nn} \end{vmatrix}$$

soit différent de zéro.

Si cette condition est réalisée, les solutions seront données par les formules de Cramer

$$\varphi_i = \frac{A_{i1}f_1 + A_{i2}f_2 + \ldots + A_{in}f_n}{\Delta},$$

en désignant, suivant l'habitude, par A_{ik} le mineur, affecté d'un signe, du terme de la $i^{\text{ième}}$ ligne et de la $k^{\text{ième}}$ colonne dans le déterminant Δ.

Cette méthode comprend maintenant un deuxième temps qui est le passage à la limite : il faut s'assurer que, lorsque n tend vers l'infini, les quantités $y_i = \varphi_i$ relatives à $x = x_i$ déterminent une courbe limite qui est solution de l'équation fonctionnelle. Hilbert [1] l'a fait lorsque le noyau K est *symétrique* et a retrouvé les résultats de Fredholm ; dans ce cas particulier, si l'on pose $\lambda = -\frac{1}{s}$, l'équation $\Delta(\lambda) = 0$ prend la forme de l'équation en s d'une quadrique ayant pour équation

$$\sum_{p,q} K_{pq} x_p x_q = 0.$$

2. C'est donc du déterminant Δ que l'on peut déduire si le problème admet une solution unique, ou s'il est impossible ou indéterminé ; c'est donc le premier membre de l'équation fonctionnelle qui règle le problème. Les racines de l'équation $\Delta(\lambda) = 0$ tendent vers les zéros du déterminant $D(\lambda)$ de Fredholm ; elles formeront donc le *spectre de valeurs propres*.

Soit la forme quadratique symétrique

$$K_{ij} x_i x_j; \tag{1}$$

à une racine λ_i de l'équation $\Delta(\lambda) = 0$ correspond une solution

$$\varphi_1^i, \quad \varphi_2^i, \quad \ldots, \quad \varphi_n^i,$$

soit Φ^i, qui est unique si λ est racine simple ; à deux valeurs distinctes λ_i et λ_j de λ correspondent des solutions *orthogonales*, c'est-à-dire telles que

$$\sum_k \varphi_k^i \varphi_k^j = 0. \tag{2}$$

Si l'équation en λ présente une racine double, il y aura plusieurs solutions qui forment une multiplicité et l'on prendra deux vecteurs Φ indé-

---

(1) *Voir* HILBERT, *Göttinger Nachrichten*, 1904.

pendants dans la multiplicité; on pourra les choisir orthogonaux (d'après Schmidt).

On obtient donc finalement un système de vecteurs Φ orthogonaux deux à deux; on peut leur donner une longueur unité, autrement dit les normer. Alors, par le changement de variables,

$$(3)\qquad \begin{cases} x_1 = \varphi_1^1 y_1 + \varphi_1^2 y_2 + \ldots + \varphi_1^n y_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ x_n = \varphi_n^1 y_1 + \varphi_n^2 y_2 + \ldots + \varphi_n^n y_n, \end{cases}$$

qui revient à prendre les Φ comme nouveaux axes, la forme quadratique devient

$$(3)\qquad \sum_i \frac{y_i^2}{\lambda_i},$$

et les anciens coefficients de la forme quadratique

$$(4)\qquad K_{ij} = \sum_k \frac{\varphi_i^k \varphi_j^k}{\lambda_k}.$$

On le vérifiera immédiatement par la transformation inverse

$$(5)\qquad y_i = \sum_h \varphi_h^i x_h.$$

Ces résultats se généralisent et s'étendent lorsque n devient infini. La forme quadratique (1) est remplacée par

$$(1')\qquad \int_a^b \int_a^b K(s,\, t)\, x(s)\, x(t)\, ds\, dt.$$

A λ_i, solution de $D(\lambda) = 0$, correspondra une fonction φ_i solution du système homogène associé (si λ_i est racine multiple, il y aura plusieurs φ_i que l'on pourra choisir orthogonales). Ces fonctions seront orthogonales deux à deux, et, après les avoir normées, on aura

$$(2')\qquad \int_a^b \varphi_i \varphi_j\, dt = \delta_{ij},$$

avec

$$\delta_{ij} = 0 \quad \text{si} \quad i \neq j,$$
$$\delta_{ij} = 1 \quad \text{si} \quad i = j.$$

Si l'on pose

$$(5')\qquad y_\alpha = \int_a^b \varphi_\alpha(t)\, x(t)\, dt,$$

on peut démontrer que (1′) deviendra

$$\sum_{1}^{\infty} \frac{y_\alpha^2}{\lambda_\alpha}, \tag{3'}$$

et que K(*s*, *t*) pourra se développer, en fonction des solutions fondamentales, par

$$K(s, t) = \sum_j \frac{\varphi_j(s)\,\varphi_j(t)}{\lambda_j}.$$

Si K n'était plus symétrique, au lieu d'envisager une forme quadratique, il faudrait envisager une forme bilinéaire qui ne peut pas toujours se ramener à une forme canonique si simple (Jordan).

III. — DEUXIÈME MÉTHODE DE HILBERT.

1. Hilbert transforme le problème de Fredholm, et le ramène à la *résolution d'un système d'une infinité d'équations linéaires à une infinité d'inconnues*. Cette méthode, reposant sur la considération de systèmes orthogonaux de fonctions, peut être suggérée par les séries de Fourier qui s'introduisent dans la résolution de l'équation $\frac{d^2 u}{ds^2} + \lambda u = 0$ dont nous avons déjà parlé. Donnons quelques explications préliminaires.

On dit qu'un système de fonctions φ_i est *orthogonal* et *normal* dans l'intervalle *ab*, si

$$\int_a^b \varphi_i(x)\,\varphi_k(x)\,dx = \delta_{ik},$$

avec

$$\delta_{ik} = 0 \quad \text{si} \quad i \neq k \ (\textit{orthogonal});$$
$$\delta_{ik} = 1 \quad \text{si} \quad i = k \ (\textit{normal}).$$

Comme exemple, nous pouvons citer le système de Fourier :

$$1, \quad \cos x, \quad \sin x, \quad \ldots, \quad \cos nx, \quad \sin nx, \quad \ldots;$$

dans l'intervalle $(0, 2\pi)$, ces fonctions sont orthogonales; il suffirait de les multiplier chacune par une constante convenable pour les normer. La notion de fonctions orthogonales est *l'extension de celle de vecteurs orthogonaux*, l'intégrale ayant remplacé la somme

$$\sum_\rho \varphi_i^\rho \varphi_k^\rho = \delta_{ik}.$$

Dans certaines conditions assez générales, une fonction f peut se développer en série de la forme

$$f(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \ldots + a_n \varphi_n(x) + \ldots,$$

ceci pouvant se rapprocher d'une égalité vectorielle, le vecteur f étant décomposé en ses composantes suivant les vecteurs $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$. Dans le cas des séries trigonométriques, Jordan a démontré que si f est une *fonction à variation bornée* (c'est-à-dire une fonction qui peut se mettre sous forme de la différence de deux fonctions croissantes et bornées), elle est égale à la série en tous les points où elle est continue (une telle fonction n'a d'ailleurs qu'une infinité dénombrable de points de discontinuité); et qu'en un point de discontinuité x, la série a pour somme $\frac{1}{2}[f(x+0)+f(x-0)]$. On prouve aisément ce résultat dans le cas particulier où f est continue et présente une dérivée. Mais Fejer a montré que même dans le cas élémentaire où f est continue, elle peut avoir une série de Fourier divergente.

Dans tous les cas, on peut se proposer *d'approcher* φ_p par une *combinaison linéaire* des n premières fonctions du système; pour mesurer l'*approximation* on peut utiliser, par exemple, le maximum de

$$|f - a_0 \varphi_0 - a_1 \varphi_1 - \ldots - a_n \varphi_n|,$$

ou bien la valeur de

$$\int_a^b (f - a_0 \varphi_0 - a_1 \varphi_1 - \ldots - a_n \varphi_n)^2 \, dx.$$

Si nous appliquons la première méthode d'approximation à l'un des cas signalés par Fejer, le maximum $\mathcal{E}_n$ tend bien vers zéro lorsque n augmente indéfiniment, mais les déterminations successives, pour les diverses valeurs de n, des coefficients $a_0, a_1, \ldots$ d'indices donnés varient et ne tendent pas vers des limites; on ne peut donc pas obtenir une série convergente de somme f.

Prenons la seconde méthode d'approximation; il faut chercher les a_i rendant l'intégrale minima; annulons les dérivées partielles par rapport aux a_i :

$$\int_a^b (f - a_0 \varphi_0 - a_1 \varphi_1 - \ldots - a_n \varphi_n) \varphi_i \, dx = 0,$$

c'est-à-dire

$$\int_a^b f \varphi_i \, dx - a_i = 0,$$

en tenant compte du fait que les fonctions φ sont orthogonales et normales. L'approximation réalisée est alors

$$\int_a^b (f - a_0\varphi_0 - a_1\varphi_1 - \ldots - a_n\varphi_n)^2\, dx$$
$$= \int_a^b f^2\, dx - 2\sum_i a_i \int_a^b f\varphi_i\, dx + 2\sum_{i\neq j} a_i a_j \int_a^b \varphi_i\varphi_j\, dx + \sum a_i^2 \int_a^b \varphi_i^2\, dx$$
$$= \int_a^b f^2\, dx - \sum a_i^2.$$

D'où la relation, due à Parseval,

$$\int_a^b f^2\, dx - \sum_{i=0}^{n} a_i^2 \geqq 0,$$

qui est vérifiée quel que soit n; donc la série Σa_i^2 est convergente et l'on a

$$\sum_{i=0}^{\infty} a_i^2 \leqq \int_a^b f^2\, dx.$$

Si l'on peut approcher arbitrairement de toute fonction f, le système de fonctions φ est dit *complet* et l'on a l'égalité

$$\int_a^b f^2(x)\, dx = \sum_{i=0}^{n} a_i^2.$$

Revenons à nos comparaisons vectorielles. $\int_a^b f^2(x)\, dx$ exprime le carré de la longueur du vecteur f; donc l'égalité de Parseval correspond au théorème de Pythagore. Plaçons-nous dans l'espace ordinaire à trois dimensions; si l'on a les projections d'un vecteur V sur trois axes perpendiculaires, on a l'égalité

$$|V|^2 = X^2 + Y^2 + Z^2;$$

si l'on ne dispose que de deux axes perpendiculaires,

$$|V|^2 \geqq X^2 + Y^2.$$

Notre système de deux axes est *incomplet*; on peut le compléter par un troisième axe perpendiculaire aux deux premiers; tandis que lorsqu'on a trois axes perpendiculaires, on ne peut en trouver un quatrième qui leur soit perpendiculaire : le système est *complet*. De même, pour un

système complet de fonctions orthogonales, les égalités d'orthogonalité

$$\int_a^b \varphi_i \Phi \, dx = 0,$$

quel que soit i, entraînent $\Phi = 0$.

2. Hilbert transforme le problème de l'équation intégrale et le ramène à la résolution d'une infinité d'équations linéaires à une infinité d'inconnues. Dans l'aperçu que nous allons donner, nous n'aborderons aucune des questions de convergence qui se présentent.

Nous supposerons que la solution φ peut se développer dans le système orthogonal φ_i de fonctions, soit

$$\varphi = \sum_{i=0}^{\infty} a_i \varphi_i,$$

avec

$$a_i = \int_a^b \varphi \varphi_i \, dx.$$

Les a_i sont dits coefficients de Fourier de φ dans le système φ_i. Si la solution n'était pas développable, il y aurait peu de choses à changer, car on pourrait l'approcher arbitrairement avec une combinaison linéaire des φ_i. Soit l'équation intégrale

$$\varphi(s) + \lambda \int_a^b K(s, t) \, \varphi(t) \, dt = f(s).$$

Nous avons, en permutant les signes $\int$ et $\sum$ sans s'inquiéter de la rigueur,

$$\lambda \int_a^b K(s, t) \, \varphi(t) \, dt = \lambda \int_a^b K(s, t) \sum_{i=0}^{\infty} a_i \varphi_i(t) \, dt = \sum_{i=0}^{\infty} \lambda a_i \int_a^b K(s, t) \, \varphi_i(t) \, dt.$$

Mais remarquons que

$$\int_a^b K(s, t) \, \varphi_i(t) \, dt = K_i(s)$$

n'est autre qu'un coefficient du développement de $K(s, t)$ considérée comme fonction de t. L'équation prend la forme

$$\varphi(s) + \lambda \sum_{i=0}^{\infty} K_i(s) \, a_i = f(s).$$

[La série obtenue est convergente, car ΣK_i^2 et Σa_i^2 le sont d'après la relation de Parseval et $K_i a_i < \frac{K_i^2 + a_i^2}{2}$.]

Nous pouvons maintenant développer les deux membres de notre relation, ou, si l'on préfère, les multiplier par φ_j et intégrer de a à b; il vient

$$(S) \qquad a_j + \lambda \sum_i k_{ij} a_i = f_j \qquad \text{pour } j = 1, 2, \ldots, n, \ldots,$$

en posant

$$k_{ij} = \int_a^b K_i(s) \varphi_j(s)\, ds,$$

qui est donc une intégrale double de $K(s, t)$ et l'on a une relation analogue à celle de Parseval :

$$\sum_{i,j} k_{ij}^2 \leqq \int\int K^2(s, t)\, ds\, dt.$$

Si l'équation intégrale a une solution φ, *les coefficients* a_i *de son développement satisfont au système linéaire* (S). *Réciproquement, si l'on a une solution* a_i *de* (S) *et si la série* Σa_i^2 *est convergente, il lui correspond une solution* φ *de l'équation intégrale.*

Dans le cas particulier où $K(s, t)$ est un *noyau dégénéré ou noyau de Goursat*, c'est-à-dire où il peut s'écrire

$$K(s, t) = S_1 T_1 + S_2 T_2 + \ldots + S_i T_i,$$

S_i étant une fonction de s et T_i une fonction de t, le système (S) se réduit à un nombre fini d'équations. Cette méthode permet de traiter un noyau quelconque par passage à la limite.

3. Nous sommes donc conduits à nous demander dans quelles conditions le système (S) présente des solutions et si les résultats obtenus pour un nombre fini d'équations à un nombre fini d'inconnues se conservent. Ces problèmes et d'autres connexes ont été étudiés par Hilbert et ses élèves (Hellinger, Toeplitz, Hilb, F. Riesz, Haar, H. Weyl, etc.); les résultats de cette étude constituent la géométrie de l'espace de Hilbert.

On a d'abord fait la théorie des formes linéaires $\sum_{p=0}^{\infty} l_p x_p$ dans le cas particulier où la série Σl_p^2 est convergente (la forme est alors dite

complètement continue) et en envisageant les valeurs des variables rendant Σx_p^2 inférieur ou égal à 1. Puis les formes quadratiques *complètement continues* (cas où la série $\Sigma k_{i,j}^2$ est convergente); on démontre que ces formes peuvent se ramener à la même expression canonique que celles à un nombre fini de variables; à cet effet on les tronque, on cherche les valeurs caractéristiques des formes tronquées et leurs limites lorsque le nombre des termes considérés augmente indéfiniment; ces limites sont les valeurs caractéristiques de la forme quadratique et par le changement linéaire correspondant de variables, on obtient la forme canonique

$$\frac{Y_1^2}{\lambda_1} + \frac{Y_2^2}{\lambda_2} + \ldots + \frac{Y_n^2}{\lambda_n} + \ldots.$$

Dans le cas de la continuité complète, il y a donc un parallélisme absolu.

Hilbert a aussi étudié les formes quadratiques $\sum_{p,q=1}^{\infty} k_{pq} x_p x_q$ dites *bornées*. Ce sont celles pour lesquelles les formes *tronquées* $\sum_{p,q=1}^{n} k_{pq} x_p x_q$ restent, quel que soit n, inférieures en module à un nombre fixe M lorsque le point $(x_1, \ldots, x_n)$ décrit la sphère de rayon 1 $(x_1^2 + x_2^2 + \ldots + x_n^2 = 1)$. Les formes complètement continues sont bornées. La réciproque n'est pas vraie. Pour les formes bornées, en général, et pour les systèmes linéaires dont les premiers membres sont leurs dérivées partielles par rapport à $x_1, x_2, \ldots$, les problèmes de valeurs propres, de représentation canonique, de rèsolution, etc., sont bien étudiés. La forme canonique $\sum_{n=1}^{\infty} \frac{y_n^2}{\lambda_n}$ ne suffit pas. Il faut y ajouter une intégrale de Stieltjes provenant du *spectre continu* de la forme, lequel n'a rien de commun avec le spectre ponctuel qui s'introduisait pour les formes complètement continues.

Parallèlement à la théorie des formes linéaires et quadratiques, à une infinité d'inconnues, on développe la théorie des *matrices associées* (matrices ayant pour éléments les coefficients de la forme) et ces deux méthodes trouveront leurs applications en physique.

Les faits nouveaux (spectre continu) révélés par l'étude des formes bornées ont eu leur répercussion dans la théorie parallèle des équations intégrales singulières où le noyau n'est plus un noyau de Fredholm.

IV. — LA GÉOMÉTRIE DE L'ESPACE DE HILBERT.

1. Sous l'influence des travaux précédents s'est développée la géométrie de l'espace de Hilbert. On peut la présenter sous la forme d'une géométrie analytique. On dira que les quantités a_1, a_2, ..., a_n, ... définissent un point P de l'espace de Hilbert si la série $\Sigma|a_i|^2$ est convergente.

E. Schmidt a généralisé les propriétés de l'espace vectoriel à l'espace de Hilbert.

F. Riesz a mis sous forme fonctionnelle la géométrie de l'espace de Hilbert en introduisant l'intégrale de Lebesgue ([1]); à toute fonction à carré sommable correspond un point de l'espace de Hilbert par l'intermédiaire d'un système orthogonal, normé et complet de fonctions. Riesz et Fischer ont démontré qu'il y avait correspondance biunivoque entre les fonctions à carré sommable et les points de l'espace de Hilbert, c'est-à-dire que réciproquement, à tout système de nombres a_i tel que la série $\Sigma|a_i|^2$ soit convergente, correspond une fonction à carré sommable définie presque partout $\left(\text{autrement dit partout, sauf, peut-être, sur un ensemble de points de mesure nulle; ce qui n'a d'ailleurs aucune importance, car la fonction n'intervient que sous le signe } \int\right)$. Comme exemple d'application, la relation entre les fonctions f et φ déterminée par

$$f(s) = \varphi(s) + \lambda \int_a^b K(s, t)\, \varphi(t)\, dt$$

correspond dans l'espace de Hilbert à la *tranformation linéaire* entre les coordonnées y et f et x de φ, définie par

$$y_j = x_j + \lambda \sum_i k_{ij} x_i,$$

soit

$$y_i = \sum_k a_{ik} x_k$$

avec Σa_{ik}^2 convergente.

On peut alors chercher les *invariants* de cette transformation, puis la *forme canonique* de celle-ci. La résolution d'une équation de Fredholm

([1]) Une fonction intégrable au sens de Lebesgue est dite *sommable*. Le carré d'une fonction sommable est sommable, mais la réciproque est fausse; le produit de deux fonctions à carrés sommables est à carré sommable (inégalité de Schwarz).

et celle d'un système linéaire à une infinité d'inconnues correspondent à l'*inversion* d'une transformation linéaire.

Sous cette double forme [*a*) géométrie analytique (point de vue algébrique des formes et des matrices, bornées ou non); *b*) résolution des équations intégrales, étude de leurs valeurs caractéristiques, etc.], le passage de l'une à l'autre pouvant s'effectuer comme l'ont montré F. Riesz et Fischer, la géométrie de l'espace de Hilbert a fait depuis trente ans de considérables progrès. Sous la seconde forme, l'étude des *équations intégrales singulières*, de leurs spectres continus, etc., pour des noyaux très généraux, a fait, entre autres, l'objet de profonds travaux de T. Carleman (voir *Sur les équations intégrales à noyau réel et symétrique*, Upsal, 1923; voir aussi *Conférence du Congrès de Zurich*, 1932, pour une vue d'ensemble). Pour les travaux de Hilbert et de ses élèves, voir notamment Hellinger-Töplitz (*Integralgleichungen und Gleichungen mit Unendlichvielen Unbekannten*, Teubner, 1928).

Cette étude trouve son application en physique dans la théorie des quanta. D'où un nouvel intérêt, qui suscite d'importants travaux; mais les représentations analytiques deviennent de plus en plus compliquées. Alors von Neumann envisage ces questions d'un point de vue différent.

2. A l'exemple de ce que, par exemple, Hilbert a fait pour l'espace euclidien ([1]), von Neumann construit axiomatiquement l'espace de Hilbert et en déduit des conséquences intéressantes sur les transformations linéaires, bornées ou non de l'espace de Hilbert. Sa méthode, très intuitive et maniable, est peut-être moins puissante que celle de Carleman; de même la géométrie élémentaire est moins puissante que la géométrie analytique, mais, sur des questions simples, elle donne des solutions moins compliquées. Dans sa voie abstraite, il semble que von Neumann ait été guidé par F. Riesz qui, dans un Mémoire intéressant des *Acta*, démontrait le théorème de Fredholm dans tous les cas où la fonctionnelle satisfait simplement à la condition

$$\mathfrak{G}(\varphi) < A\,|\varphi|,$$

A désignant une constante.

([1]) *Voir* Hilbert, *Grundlagen der Geometrie*. Comme précurseurs il avait eu Lie, Helmholtz et même Euclide.

V. — PROGRAMME DE CES LEÇONS.

Voici maintenant le *programme* que nous allons suivre :

a. *Les transformations linéaires dans l'espace à n dimensions;* nous étudierons parallèlement la résolution des équations linéaires, le calcul des matrices, la réduction des formes quadratiques et hermitiennes; en passant, les équations linéaires seront étudiées sans déterminants par une méthode qui peut se généraliser à une infinité d'inconnues; les valeurs caractéristiques d'une forme quadratique seront caractérisées par des propriétés d'extremum, en les rattachant à l'équation en s relative à la recherche des directions principales d'une quadrique : le plus grand λ (plus grand axe) correspond au maximum de la forme quadratique, le λ immédiatement inférieur au plus petit des maxima des rayons vecteurs d'une section plane à $n-1$ dimensions passant par le grand axe — autrement dit satisfaisant à une relation linéaire — le suivant au plus petit des maxima des rayons vecteurs d'une section plane à $n-2$ dimensions satisfaisant à deux relations linéaires, et ainsi de suite...

b. Nous étendrons à l'espace de Hilbert à une infinité de dimensions les résultats obtenus sur les équations linéaires; puis nous étudierons les formes linéaires, bilinéaires et quadratiques complètement continues, les formes bilinéaires, quadratiques, hermitiennes bornées, et parallèlement les matrices correspondantes. C'est une étude algébrique des transformations ou opérateurs linéaires de l'espace de Hilbert.

c. Parallèlement à *b*, nous étudierons les équations intégrales régulières correspondant à des *noyaux bornés*.

d. Nous étudierons la théorie axiomatique des espaces de Hilbert; les opérateurs hermitiens bornés ou non; les équations intégrales régulières ou singulières.

En somme, après un préambule *a*, nous présenterons parallèlement, en *b*, *c*, *d*, sous sa double forme, *analytique* (algébrique ou fonctionnelle) ou *axiomatique*, l'étude, que nous allons entreprendre, des *transformations* ou *opérateurs* linéaires dans *l'espace de Hilbert*.

CHAPITRE I.

I. — RAPPEL DE RÉSULTATS CONCERNANT LA RÉSOLUTION ET LA DISCUSSION DES SYSTÈMES D'ÉQUATIONS LINÉAIRES PAR LA MÉTHODE DES DÉTERMINANTS.

1. Nous allons d'abord rappeler rapidement la théorie classique de la résolution d'un système linéaire, en utilisant les déterminants; nous montrerons ensuite que l'on peut faire disparaître ceux-ci des résultats, et enfin nous donnerons une autre méthode d'étude, indépendante des déterminants, qui sera susceptible de se généraliser au cas d'un nombre infini d'équations.

Soit d'abord un système de n équations à n inconnues :

$$(\mathrm{N}) \qquad \left\{ \begin{array}{l} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = y_1, \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{n1} x_1 + a_{n2} x_2 + \ldots + a_{nn} x_n = y_n, \end{array} \right.$$

et nous supposerons, en premier lieu, que le déterminant formé par les coefficients des inconnues n'est pas nul :

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} = |\,a_{ik}\,| \neq 0.$$

Nous désignerons par A_{ik} le *complément algébrique* ou *mineur* affecté d'un signe $[(-1)^{i+k}]$ du terme a_{ik} dans le déterminant Δ; A_{ik} sera, si l'on veut, défini par la relation

$$\mathrm{A}_{ik} = \frac{\partial \Delta}{\partial a_{ik}}.$$

D'après la règle du développement d'un déterminant par rapport aux

éléments d'une colonne, nous aurons

$$\Delta = A_{1j}a_{1j} + A_{2j}a_{2j} + \ldots + A_{nj}a_{nj} = \sum_{i=1}^{n} A_{ij}a_{ij}.$$

Si, d'autre part, nous considérons la somme

$$A_{1j}a_{1k} + A_{2j}a_{2k} + \ldots + A_{nj}a_{nk} = \sum_{i=1}^{n} A_{ij}a_{ik},$$

dans laquelle j est différent de k, elle représente le développement d'un déterminant qui se réduit de Δ en remplaçant la $j^{\text{ième}}$ colonne par la $k^{\text{ième}}$; ce nouveau déterminant ayant deux colonnes identiques est nul; nous pouvons donc écrire d'une façon générale

$$\sum_{i} A_{ij}a_{ik} = \delta_{jk}\Delta.$$

Multiplions alors la première équation du système (N) par A_{1k}, la deuxième par A_{2k}, ..., la $n^{\text{ième}}$ par A_{nk} et ajoutons; il vient, en tenant compte de la relation que nous venons d'établir,

$$\Delta x_k = \sum_{i} A_{ik}y_i,$$

soit, puisque Δ n'est pas nul,

$$x_k = \sum_{i} \frac{A_{ik}}{\Delta} y_i.$$

Nous obtenons ainsi les formules, dites de Cramer, qui nous donnent, *en supposant son existence*, la solution du système (N). Pour être rigoureux, il faut reporter ces valeurs dans l'une quelconque des équations et montrer qu'elle est satisfaite; par exemple prenons la première

$$\sum_{k} a_{1k} \sum_{i} \frac{A_{ik}}{\Delta} y_i = y_1,$$

ou bien

$$\sum_{i} y_i \sum_{k} \frac{a_{1k}A_{ik}}{\Delta} = y_1,$$

$$\sum_{i} y_i \delta_{i1} = y_1;$$

elle est bien vérifiée.

Nous avons donc démontré que, si Δ est différent de zéro, le système (N) a une solution et une seule *quels que soient les seconds*

membres. Cette solution est linéaire par rapport à ceux-ci; attribuons-leur maintenant des valeurs particulières :

$$Y^{(1)},\quad \text{soit}\quad y_1^{(1)} = 1,\quad y_2^{(1)} = 0,\quad y_3^{(1)} = 0,\quad \ldots,\quad y_n^{(1)} = 0,$$

qui donne la solution

$$X^{(1)},\quad \text{soit}\quad x_1^{(1)} = \frac{A_{11}}{\Delta},\quad x_2^{(1)} = \frac{A_{12}}{\Delta},\quad x_3^{(1)} = \frac{A_{13}}{\Delta},\quad \ldots,\quad x_n^{(1)} = \frac{A_{1n}}{\Delta};$$

$$Y^{(2)},\quad \text{soit}\quad y_1^{(2)} = 0,\quad y_2^{(2)} = 1,\quad y_3^{(2)} = 0,\quad \ldots,\quad y_n^{(2)} = 0,$$

qui donne la solution

$$X^{(2)},\quad \text{soit}\quad x_1^{(2)} = \frac{A_{21}}{\Delta},\quad x_2^{(2)} = \frac{A_{22}}{\Delta},\quad x_3^{(2)} = \frac{A_{23}}{\Delta},\quad \ldots,\quad x_n^{(2)} = \frac{A_{2n}}{\Delta};$$

de même $Y^{(3)}$, $Y^{(4)}$, et enfin

$$Y^{(n)},\quad \text{soit}\quad y_1^{(n)} = 0,\quad y_2^{(n)} = 0,\quad y_3^{(n)} = 0,\quad \ldots,\quad y_{n-1}^{(n)} = 0,\quad y_n^{(n)} = 1,$$

qui donne la solution

$$X^{(n)},\quad \text{soit}\quad x_1^{(n)} = \frac{A_{n1}}{\Delta},\quad x_2^{(n)} = \frac{A_{n2}}{\Delta},\quad x_3^{(n)} = \frac{A_{n3}}{\Delta},\quad \ldots,$$
$$x_{n-1}^{(n)} = \frac{A_{n,n-1}}{\Delta},\quad x_n^{(n)} = \frac{A_{nn}}{\Delta}.$$

Avec ces solutions particulières, que nous appellerons *solutions élémentaires*, nous formerons immédiatement la solution générale relative à un second membre quelconque

$$Y,\quad \text{soit}\quad y_1,\quad y_2,\quad \ldots,\quad y_n,$$

que nous écrirons symboliquement

$$Y = y_1 Y^{(1)} + y_2 Y^{(2)} + \ldots + y_n Y^{(n)}.$$

Cette solution sera

$$X,\quad \text{soit}\quad x_1 = \sum_i x_1^{(i)} y_i,\qquad x_2 = \sum_i x_2^{(i)} y_i,\qquad \ldots,\qquad x_n = \sum_i x_n^{(i)} y_i,$$

que nous écrirons symboliquement

$$X = y_1 X^{(1)} + y_2 X^{(2)} + \ldots + y_n X^{(n)}.$$

Remarquons que c'est là une *forme classique de solution générale dans un problème linéaire*; par exemple : recherche d'une solution de l'équation $\Delta U = 0$ qui prenne sur des surfaces données des valeurs cons-

tantes données (problème de Dirichlet en physique, problème de l'équilibre électrique). On prend pour constantes successivement 1 sur chaque conducteur et 0 sur tous les autres et l'on obtient ainsi les solutions élémentaires; la solution cherchée est

$$U = \sum_i y_i U^{(i)},$$

les y_i étant les constantes données (ce sont les potentiels dans le cas de l'équilibre électrique).

Remarquons encore que si nous donnons une valeur nulle aux seconds membres, les formules de Cramer nous montrent que le système présente uniquement la solution 0.

2. On dit que r formes linéaires $f_1, f_2, \ldots, f_r$ sont *linéairement dépendantes* si l'on peut trouver r constantes $\lambda_1, \lambda_2, \ldots, \lambda_r$, *non toutes nulles*, telles que

$$\lambda_1 f_1 + \lambda_2 f_2 + \ldots + \lambda_r f_r \equiv 0. \tag{1}$$

Dans le cas contraire, les r formes sont dites *linéairement indépendantes*.

Nous allons montrer que, quand Δ n'est pas nul, les n formes constituant les premiers membres de (N) sont linéairement indépendantes; en effet, développons la condition (1) :

$$(\mathrm{H}') \quad \begin{cases} \lambda_1 a_{11} + \lambda_2 a_{21} + \ldots + \lambda_n a_{n1} = 0, \\ \lambda_1 a_{12} + \lambda_2 a_{22} + \ldots + \lambda_n a_{n2} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ \lambda_1 a_{1n} + \lambda_2 a_{2n} + \ldots + \lambda_n a_{nn} = 0. \end{cases}$$

Nous obtenons ainsi un système homogène (H') dont le tableau des coefficients des inconnues se déduit de celui du système (N) par permutation des lignes et des colonnes; (H') est dit *système homogène transposé* de (N). Son déterminant est égal à Δ (car on ne change pas la valeur d'un déterminant en permutant lignes et colonnes), donc différent de zéro et, d'après une remarque déjà faite, ce système homogène ne présente que zéro comme solutions :

$$\lambda_i = 0.$$

Les formes sont linéairement indépendantes.

Nous venons de démontrer qu'une condition suffisante pour que n formes linéaires à n variables soient linéairement indépendantes est que

le déterminant de leurs coefficients ne soit pas nul; nous verrons qu'elle est nécessaire.

3. Nous nous plaçons toujours dans le cas d'un système de n équations à n inconnues, mais nous supposons maintenant que le déterminant Δ est nul. Considérons, parmi les déterminants extraits du déterminant Δ, ceux qui ne sont pas nuls; à moins que les coefficients a_{ik} ne soient tous nuls, ces déterminants existent (pour s'en assurer, il suffit de prendre ceux d'ordre 1 réduit à un coefficient non nul); les ordres de ces déterminants sont tous inférieurs à n, ils présentent donc une borne supérieure ρ qui est inférieure à n. Le nombre entier ρ est appelé le *rang* du système. En changeant les indices des formes et les indices des variables, nous pouvons toujours supposer que le déterminant

$$\delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1\rho} \\ a_{21} & a_{22} & \dots & a_{2\rho} \\ \dots & \dots & \dots & \dots \\ a_{\rho 1} & a_{\rho 2} & \dots & a_{\rho\rho} \end{vmatrix}$$

n'est pas nul; nous l'appellerons *déterminant principal*, les ρ premières équations seront les *équations principales*, et les ρ premières inconnues les *inconnues principales*. Ceci entraîne comme conséquence que les formes $f_1, f_2, \dots, f_\rho$ sont linéairement indépendantes, car dans le système en λ on obtient la solution zéro (en ne tenant compte que de l'identification relative aux ρ premières variables). Montrons maintenant que chacune des autres formes est une combinaison linéaire des ρ premières; en effet, le déterminant

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1\rho} & a_{1\rho+h} \\ a_{21} & a_{22} & \dots & a_{2\rho} & a_{2\rho+h} \\ \dots & \dots & \dots & \dots & \dots\dots \\ a_{\rho 1} & a_{\rho 2} & \dots & a_{\rho\rho} & a_{\rho\rho+h} \\ a_{\rho+1,1} & a_{\rho+1,2} & \dots & a_{\rho+1,\rho} & a_{\rho+1,\rho+h} \end{vmatrix}$$

est nul, si h est négatif ou nul parce qu'il a deux colonnes identiques, si h est positif en raison de la définition même du rang ρ. Développons ce déterminant par rapport aux éléments de la dernière colonne

$$\lambda_1 a_{1,\rho+h} + \lambda_2 a_{2,\rho+h} + \dots + \lambda_{\rho+1} a_{\rho+1,\rho+h} = 0.$$

Les λ s'obtiennent à l'aide des ρ premières colonnes, donc ne dépendent pas de h; d'où la relation

$$\lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_\rho f_\rho + \lambda_{\rho+1} f_{\rho+1} \equiv 0,$$

car là relation obtenue plus haut entre les a montre que chaque inconnue x a son coefficient nul. Mais $\lambda_{\rho+1} = \delta \neq 0$; nous pouvons donc écrire

$$f_{\rho+1} \equiv \alpha_{11} f_1 + \alpha_{12} f_2 + \ldots + \alpha_{1\rho} f_\rho.$$

Nous obtiendrions de même

$$f_{\rho+l} \equiv \alpha_{l1} f_1 + \alpha_{l2} f_2 + \ldots + \alpha_{l\rho} f_\rho \qquad \text{avec} \quad l > 0.$$

Nous pouvons donc dire que le *rang* ρ du système est le *nombre maximum de formes linéairement indépendantes* choisies parmi les f_i. (C'est un exemple de définition dans laquelle nous avons fait disparaître l'expression : déterminant.)

La condition nécessaire et suffisante pour que les $n - \rho$ dernières équations (N) soient vérifiées lorsque les ρ premières le sont est qu'entre les seconds membres existent les relations

$$y_{\rho+l} = \alpha_{l1} y_1 + \alpha_{l2} y_2 + \ldots + \alpha_{l\rho} y_\rho.$$

Dans la théorie avec déterminants on met ces équations sous la forme

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1\rho} & y_1 \\ a_{21} & a_{22} & \ldots & a_{2\rho} & y_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{\rho 1} & a_{\rho 2} & \ldots & a_{\rho\rho} & y_\rho \\ a_{\rho+l,1} & a_{\rho+l,2} & \ldots & a_{\rho+l,\rho} & y_{\rho+l} \end{vmatrix} = 0.$$

Les premiers membres de ces relations sont les *déterminants caractéristiques*. Si les seconds membres du système (N) vérifient ces relations, le système se réduit à ses équations principales; sinon le système est impossible.

Prenons le premier cas : dans les équations principales nous pouvons choisir arbitrairement les valeurs des inconnues $x_{\rho+1}, x_{\rho+2}, \ldots, x_n$; le système des équations principales peut alors se résoudre par rapport aux inconnues principales par les formules de Cramer. Le problème est ici indéterminé, la solution *dépend de $n - \rho$ arbitraires*.

Outre le système considéré (N), nous allons être amenés à considérer le système homogène :

$$\text{(H)} \qquad \begin{cases} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = 0, \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{n1} x_1 + a_{n2} x_2 + \ldots + a_{nn} x_n = 0, \end{cases}$$

et le système homogène transposé :

$$(H') \quad \begin{cases} a_{11}\lambda_1 + a_{21}\lambda_2 + \ldots + a_{n1}\lambda_n = 0, \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \ldots + a_{n2}\lambda_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{1n}\lambda_1 + a_{2n}\lambda_2 + \ldots + a_{nn}\lambda_n = 0. \end{cases}$$

Ces trois systèmes ont les mêmes déterminants, extraits du tableau des coefficients (pour H′, il suffit de permuter lignes et colonnes, ce qui ne change pas la valeur du déterminant). Donc (H′) est de rang ρ; les conditions relatives aux déterminants caractéristiques sont satisfaites puisque tous les seconds membres sont nuls; nous pouvons choisir arbitrairement $\lambda_{\rho+1}$, $\lambda_{\rho+2}$, ..., λ_n, et résoudre les ρ premières équations par rapport à λ_1, λ_2, ..., λ_ρ; les $n-\rho$ dernières équations sont vérifiées. Les solutions du système (H′) dépendent donc de $n-\rho$ constantes arbitraires $\lambda_{\rho+1}$, ..., λ_n; on peut introduire un *système fondamental* formé de $n-\rho$ solutions indépendantes en donnant successivement à chacun de ces λ la valeur 1 et aux autres la valeur 0.

De même les solutions du système (H) dépendent aussi de $n-\rho$ constantes arbitraires; la solution générale du système (N) peut d'ailleurs se mettre sous la forme de la somme d'une solution particulière et de la solution générale du système H.

Finalement, nous arrivons à l'énoncé du *théorème général de l'alternative :*

1° *Ou bien le système* (N) *admet une solution quels que soient les* y_i [$\Delta \neq 0$], *alors la solution est unique et les systèmes* (H) *et* (H′) *n'admettent que zéro pour solution* (ces trois affirmations sont équivalentes);

2° *Ou bien le système* (H) *et le système* (H′) *ont des solutions différentes de zéro; ils ont le même nombre de solutions linéairement indépendantes; soit r ce nombre* [$\Delta = 0$; rang $\rho = n - r$]; *le système* (N) *n'est résoluble que pour des seconds membres* y_i *particuliers satisfaisant aux r relations qui existent alors entre les formes des premiers membres; et s'il est résoluble, la solution dépend de r constantes arbitraires* (*correspondant à la solution générale du système homogène*).

Dans cet énoncé on peut supprimer ce qui concerne les déterminants, et qui se trouve entre crochets. Ce *théorème de l'alternative* peut être obtenu sans déterminants et peut se *généraliser à des problèmes linéaires non algébriques.*

4. Disons quelques mots sur la résolution d'un système linéaire de m équations à n inconnues :

$$\begin{aligned}
a_{11}\ x_1 + a_{12}\ x_2 + \ldots + a_{1n}\ x_n &= y_1,\\
a_{21}\ x_1 + a_{22}\ x_2 + \ldots + a_{2n}\ x_n &= y_2,\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots&,\\
a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n &= y_m.
\end{aligned}$$

En recherchant le nombre maximum de formes linéairement indépendantes, nous obtenons le rang du système

$$\rho \leqq m,$$
$$\rho \leqq n.$$

Pour cette recherche, nous pourrons utiliser les déterminants : il y a un déterminant principal, des équations principales que nous écrirons les premières et des inconnues principales que nous affecterons aussi des plus petits indices.

a. Si $m < n$, distinguons deux cas :

1° $\rho = m$. — Nous choisirons arbitrairement les valeurs des inconnues $x_{m+1}, x_{m+2}, \ldots, x_n$ et résoudrons par rapport aux inconnues principales ; le système sera donc toujours possible et présentera une infinité de solutions dépendant de $n - \rho$ constantes arbitraires.

2° $\rho < m$. — Les formes ne sont plus indépendantes. Si les seconds membres annulent les déterminants caractéristiques, il y a une infinité de solutions dépendant de $n - \rho$ constantes arbitraires. Sinon, le système est impossible.

b. Si $m > n$, alors $\rho \leqq n < m$. — Les formes ne sont pas indépendantes (¹) et comme dans le dernier cas il faut former les caractéristiques.

En résumé :

Si le rang est égal au nombre des équations, le problème est toujours possible.

S'il est de plus égal au nombre des variables, la solution est unique.

Avant de reprendre directement ces questions en vue d'une généralisation aux systèmes infinis, rappelons quelques notions relatives aux vecteurs pour simplifier notre exposé.

(¹) En particulier $n + 1$ formes linéaires à n variables ne sont jamais indépendantes.

II. — NOTIONS ÉLÉMENTAIRES DE GÉOMÉTRIE LINÉAIRE HOMOGÈNE (GÉOMÉTRIE AFFINE).

1. **Définition par le calcul.** — Un vecteur X de l'espace E_n est un ensemble de n nombres complexes rangés dans un certain ordre :

$$X = (x_1, x_2, \ldots, x_n).$$

Ces nombres sont appelés les composantes du vecteur.

Deux vecteurs sont égaux si leurs composantes sont égales :

$$X = Y$$

est équivalent à

$$x_i = y_i \quad \text{pour} \quad i = 1, 2, \ldots, n.$$

Nous allons définir sur ces vecteurs les opérations fondamentales :

a. Produit d'un vecteur par un nombre. — Le vecteur aX a pour composantes $ax_1, ax_2, \ldots, ax_n$.

b. Somme de deux vecteurs. — Le vecteur $X + Y$ a pour composantes $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n$.

De ces définitions il résulte immédiatement que la somme de plusieurs vecteurs est commutative, associative; qu'étant donné deux vecteurs X et Y, il en existe un troisième et un seulement qui ajouté au second donne le premier (différence); que le produit d'un vecteur par l'unité ne le modifie pas; le produit est distributif par rapport à l'addition des nombres et par rapport à l'addition des vecteurs; le produit est associatif par rapport aux nombres.

Cette définition par le calcul introduit un système de vecteurs coordonnés qui sont privilégiés dans E_n; ce sont les vecteurs dont une composante est égale à l'unité, toutes les autres étant nulles. Nous allons donner une définition axiomatique de l'espace vectoriel à n dimensions qui ne fera plus intervenir de vecteurs coordonnés fondamentaux.

2. **Définition axiomatique.** — Nous allons donner d'abord deux axiomes fondamentaux, qui vont nous définir des *êtres* X, Y, ...

a. Addition. — A deux de ces *êtres* on peut en associer un troisième, qu'on appelle leur somme et qui jouit des propriétés suivantes :

1° $X + Y = Y + X$ (commutabilité),

2° $(X + Y) + Z = X + (Y + Z)$ (associativité);

3° Si l'on se donne X et Y, il existe un être et un seul Z tel que $X = Y + Z$; on l'appelle différence de X et de Y et l'on écrit

$$Z = X - Y.$$

Comme conséquence il existe un *être* o qui est la différence $X - X$ et l'on a

$$X + 0 = 0 + X = X,$$

et à chaque X on peut faire correspondre un *être* opposé qui, ajouté au premier, donne zéro.

b. Multiplication. — A un de ces *êtres* X et à un nombre a, on peut en associer un second, qu'on appelle produit de l'*être* par le nombre, qu'on désigne par la notation aX et qui jouit des propriétés suivantes :

1° $$1 \times X = X,$$

2° $$(a + b)X = aX + bX$$

(distributivité par rapport à l'addition des nombres),

3° $$a(bX) = (ab)X$$

(associativité relativement aux nombres),

4° $$a(X + Y) = aX + aY$$

(distributivité par rapport à l'addition des *êtres*).

Ces *êtres* s'appellent des *vecteurs*; pour définir complètement l'espace vectoriel à n dimensions, il faut à ces deux axiomes en ajouter un troisième (axiome de la dimension); la géométrie homogène linéaire est basée sur l'ensemble de ces trois axiomes.

Avant d'introduire le troisième axiome, tirons des conséquences des deux premiers.

3. h vecteurs $X_1, X_2, \ldots, X_h$ sont *linéairement indépendants* s'il est impossible d'avoir la relation

$$a_1 X_1 + a_2 X_2 + \ldots + a_h X_h = 0,$$

à moins que tous les a ne soient nuls.

Considérons h vecteurs $X_1, X_2, \ldots, X_h$, linéairement indépendants; et soit

$$X = \xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_h X_h;$$

l'ensemble des vecteurs X, pour toutes les valeurs possibles des ξ, forme,

par définition, une *multiplicité linéaire* E_h *à h dimensions* définie par les vecteurs $X_1, X_2, \ldots, X_h$. On dit que ces vecteurs forment une *base* de la multiplicité et que $\xi_1, \xi_2, \ldots, \xi_h$ sont les *composantes* de X dans E_h relativement à cette base. Ces composantes sont *uniques;* supposons en effet que nous ayons

$$X = \xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_h X_h,$$
$$X = \xi'_1 X_1 + \xi'_2 X_2 + \ldots + \xi'_h X_h.$$

Soustrayons membre à membre :

$$(\xi_1 - \xi'_1) X_1 + (\xi_2 - \xi'_2) X_2 + \ldots + (\xi_h - \xi'_h) X_h = 0,$$

et puisque les vecteurs $X_1, X_2, \ldots, X_h$ sont linéairement indépendants, cette relation entraîne

$$\xi_i = \xi'_i \qquad \text{pour} \quad i = 1, 2, \ldots, h.$$

Tout vecteur de E_h s'exprime donc, et d'une seule façon, comme combinaison linéaire des vecteurs de base.

Montrons que $h+1$ *vecteurs de la multiplicité linéaire* E_h *ne peuvent être linéairement indépendants*. Ces vecteurs sont définis par

$$Y_i = \xi_{i1} X_1 + \xi_{i2} X_2 + \ldots + \xi_{ih} X_h \qquad \text{pour} \quad i = 1, 2, \ldots, h+1.$$

Cherchons si nous pouvons avoir une relation :

$$\lambda_1 Y_1 + \lambda_2 Y_2 + \ldots + \lambda_{h+1} Y_{h+1} = 0.$$

Soient

$$\sum_{i=1}^{h+1} \lambda_i \sum_{k=1}^{h} \xi_{ik} X_k = 0,$$
$$\sum_{k=1}^{h} X_k \sum_{i=1}^{h+1} \xi_{ik} \lambda_i = 0.$$

Les X_k étant linéairement indépendants, cette relation exige

$$\sum_{i=1}^{h+1} \lambda_i \xi_{ik} = 0 \qquad \text{pour} \quad k = 1, 2, \ldots, h$$

ou, sous la forme développée :

$$\xi_{11} \lambda_1 + \xi_{21} \lambda_2 + \ldots + \xi_{h+1,1} \lambda_{h+1} = 0,$$
$$\xi_{12} \lambda_1 + \xi_{22} \lambda_2 + \ldots + \xi_{h+1,2} \lambda_{h+1} = 0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$
$$\xi_{1h} \lambda_1 + \xi_{2h} \lambda_2 + \ldots + \xi_{h+1,h} \lambda_{h+1} = 0,$$

système homogène de h équations à $h+1$ inconnues $\lambda_1, \lambda_2, \ldots, \lambda_{h+1}$. Ce système est toujours possible ([1]) et présente des solutions non toutes nulles; la proposition est donc établie. On peut d'ailleurs trouver effectivement dans E_h, h vecteurs linéairement indépendants : il suffit de prendre $X_1, X_2, \ldots, X_h$. Il en résulte que *h est le nombre maximum de vecteurs linéairement indépendants dans* E_h.

4. Changement de base. — Considérons h vecteurs $Y_1, Y_2, \ldots, Y_h$ de E_h, définis par

$$Y_i = \sum_{j=1}^{h} c_{ij} X_j$$

et supposons que le déterminant $|c_{ij}|$ ne soit pas nul. Si nous considérons ces relations comme des équations entre les lettres X et Y, ce système peut être résolu en X par des formules telles que

$$X_j = \sum_{l=1}^{h} \gamma_{jl} Y_l.$$

Ce sont les expressions des vecteurs X en fonction des vecteurs Y; si nous les reportons dans le premier système, nous obtenons des identités.

Soit un vecteur V de la multiplicité E_h :

$$V = \xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_h X_h = \sum_{j=1}^{h} \xi_j X_j.$$

Nous avons donc

$$V = \sum_{j=1}^{h} \xi_j \sum_{l=1}^{h} \gamma_{jl} Y_l = \sum_{l=1}^{h} Y_l \sum_{j=1}^{h} \xi_j \gamma_{jl},$$

soit

$$V = \eta_1 Y_1 + \eta_2 Y_2 + \ldots + \eta_h Y_h$$

en posant

$$\eta_l = \sum_{j=1}^{h} \xi_j \gamma_{jl}.$$

Et, de même, toute combinaison linéaire des Y_i est une combinaison linéaire des X_i. Montrons que les *vecteurs Y_i sont linéairement indé-*

([1]) Nous le savons ici par le paragraphe I relatif aux déterminants. Nous l'établirons au paragraphe III sans les déterminants.

pendants; à cet effet, cherchons les λ tels que

$$\sum_i \lambda_i Y_i = 0,$$

soit

$$0 = \sum_i \lambda_i \sum_j c_{ij} X_j = \sum_j X_j \sum_i \lambda_i c_{ij},$$

et puisque les X sont linéairement indépendants,

$$\sum_{i=1}^{h} \lambda_i c_{ij} = 0 \qquad \text{pour} \quad j = 1, 2, \ldots, h.$$

C'est un système homogène de h équations à h inconnues à déterminant non nul; il ne présente que la solution $\lambda_i = 0$; les Y_i sont linéairement indépendants.

L'ensemble des vecteurs Y_i peut donc être pris comme base de la multiplicité E_h, au même titre que l'ensemble des vecteurs X_i.

Ces deux bases sont *équivalentes*. Le nombre des vecteurs d'une base quelconque de E_h est toujours h. Tout système de h vecteurs de E_h indépendants est une base. Deux bases distinctes se déduisent l'une de l'autre par une substitution linéaire à déterminant $\neq 0$.

5. **Axiome de la dimension.** — L'espace E_n lui-même possède n vecteurs linéairement indépendants qui sont (espace vectoriel à n dimensions défini par le calcul) :

$$\begin{array}{l} e_1(1, 0, 0, \ldots, 0), \\ e_2(0, 1, 0, \ldots, 0), \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ e_n(0, 0, 0, \ldots, 1), \end{array}$$

car on ne peut avoir

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n = 0,$$

à moins que les λ ne soient tous nuls. Tout vecteur X de E_n, ayant pour coordonnées $x_1, x_2, \ldots, x_n$, peut se mettre sous la forme

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n.$$

Les x_i sont les *coordonnées absolues* de X.

Ainsi que nous l'avons déjà fait remarquer, la définition de E_n par des *coordonnées absolues* introduit un système de vecteurs privilégiés, alors qu'un système quelconque de n vecteurs linéairement indépendants peut très bien être pris comme système de coordonnées.

Ceci ne se présentera plus au point de vue axiomatique. Les deux axiomes que nous avons énoncés dans le paragraphe 2 ne font pas intervenir le nombre de dimensions de l'espace vectoriel dans lequel on se place; aussi leur adjoindrons-nous *l'axiome de la dimension suivant :*

Dans l'espace vectoriel à n dimensions E_n *il existe n vecteurs linéairement indépendants et n au plus.*

Nous allons voir qu'à partir de ces trois axiomes on peut bâtir la géométrie vectorielle affine, car on retrouve la définition analytique. Soient, en effet, $e_1, e_2, \ldots, e_n$, n vecteurs linéairement indépendants dans E_n et X un vecteur quelconque de E_n; ces $n+1$ vecteurs ne peuvent être linéairement indépendants. Nous avons donc :

$$aX + \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n = 0$$

avec $a, \xi_1, \xi_2, \ldots, \xi_n$, non tous nuls. Puisque les $e_1, e_2, \ldots, e_n$ sont linéairement indépendants, a n'est pas nul; nous pouvons donc résoudre en X la relation précédente :

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n.$$

C'est la définition analytique de E_n; mais, par le procédé axiomatique, tous les systèmes de coordonnées sont *a priori* équivalents. L'application des axiomes a et b donne ensuite les règles de calcul prises comme définition au n° 1 qui précède.

6. Nous pouvons définir (par voie axiomatique toujours) une multiplicité linéaire par un système de vecteurs $X_1, X_2, \ldots, X_k, \ldots$. Les vecteurs

$$X = c_1 X_1 + c_2 X_2 + \ldots + c_l X_l$$

(c_i nombres complexes arbitraires), combinaisons linéaires d'un nombre fini (quelconque) des vecteurs X_k constituent la multiplicité linéaire définie par les vecteurs donnés, *qui peuvent être en nombre fini ou infini.* Les vecteurs donnés sont dans un espace à n dimensions; le nombre maximum d'entre eux, qui sont linéairement indépendants, est donc $h \leq n$; soient $X_1, X_2, \ldots, X_h$ de tels vecteurs et soit X un vecteur quelconque de la multiplicité. Ces $h+1$ vecteurs ne peuvent être linéairement indépendants; nous avons donc

$$aX + \xi'_1 X_1 + \xi'_2 X_2 + \ldots + \xi'_h X_h = 0$$

avec $a, \xi'_1, \xi'_2, \ldots, \xi'_h$ non tous nuls; a n'est pas nul puique les $X_1, X_2, \ldots, X_h$

sont linéairement indépendants. Nous pouvons donc résoudre en X la relation précédente

$$X = \xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_h X_h.$$

Nous retrouvons la définition d'une multiplicité linéaire à h dimensions ayant pour vecteurs de base $X_1, X_2, \ldots, X_h$.

7. **Décomposition d'un espace à n dimensions.** — Soient deux espaces partiels (ou multiplicités linéaires) E' à n' dimensions et E'' à n'' dimensions, tels que *tout vecteur de* E *puisse d'une manière et d'une seule, se décomposer en la somme de deux vecteurs, l'un appartenant à* E', *l'autre à* E''. Dans ces conditions nous dirons avoir *décomposé l'espace* E en deux sous-espaces E' et E'', et nous écrirons

$$E = E' + E'',$$

E' et E'' sont dits *complémentaires.*

Les deux sous-espaces ne peuvent avoir en commun que o, car si le vecteur $X \neq 0$ leur appartenait à tous deux, le vecteur X de E pourrait se décomposer de *deux* manières, à savoir : X (de E') + o (de E''), ou bien : o (de E') + X (de E''). Dans le sous-espace E' à n' dimensions, choisissons n' vecteurs linéairement indépendants formant un système de base : $X'_1, X'_2, \ldots, X'_{n'}$, et faisons de même dans E'' : $X''_1, X''_2, \ldots X''_{n''}$. D'après la définition même de la décomposition, X étant un vecteur quelconque de E, nous avons

$$X = X' + X'',$$

X' appartenant à E' et X'' à E''; soit

$$X = \xi'_1 X'_1 + \xi'_2 X'_2 + \ldots + \xi'_{n'} X'_{n'} + \xi''_1 X''_1 + \xi''_2 X''_2 + \ldots + \xi''_{n''} X''_{n''}.$$

Les vecteurs X'_i et X''_j sont linéairement indépendants, car si l'on avait

$$\alpha'_1 X'_1 + \alpha'_2 X'_2 + \ldots + \alpha'_{n'} X'_{n'} + \alpha''_1 X''_1 + \alpha''_2 X''_2 + \ldots + \alpha''_{n''} X''_{n''} = 0,$$

l'ensemble des n' premiers termes constituant un vecteur de E' et l'ensemble des n'' derniers un vecteur de E'', chacun de ces ensembles serait nul, E' et E'' n'ayant que o en commun. Les X' étant indépendants, ainsi que les X'', il s'ensuit la nullité de tous les α.

L'ensemble des vecteurs X'_i et X''_j constitue donc un système de base pour E qui a par conséquent $n = n' + n''$ dimensions. Donc la réunion d'un système de coordonnées de E' et d'un système de coordonnées de E'', forme un système de coordonnées de E que nous dirons être *adapté à la décomposition* $E = E' + E''$.

Nous pouvons dire encore que, si l'on se donne dans E_n deux sous-espaces n'ayant en commun que zéro, E' à n' dimensions et E'' à n'' dimensions, tels que $n = n' + n''$, l'espace E est décomposé en les deux sous-espaces E' et E''. C'est immédiat : choisissons dans E', n' vecteurs de base $X_1, X_2, \ldots, X_{n'}$, et dans E'' n'' vecteurs de base $X_{n'+1}, X_{n'+2}, \ldots, X_n$. Par une démonstration identique à la précédente, nous montrerions que ces $n = n' + n''$ vecteurs de E sont linéairement indépendants; ils forment donc une base de E. Un vecteur de E peut, d'une manière et d'une seule, se mettre sous la forme

$$X = (\xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_{n'} X_{n'}) + (\xi_{n'+1} X_{n'+1} + \ldots + \xi_n X_n),$$
$$X = X' + X''.$$

Deux *espaces complémentaires* dans E_n sont donc caractérisés par deux propriétés : ils n'ont en commun que zéro et la somme de leurs nombres de dimensions est n.

8. **Projection.** — Soit dans l'espace E à n dimensions un sous-espace E' à $n' < n$ dimensions. Deux vecteurs X et Y dont la différence appartient à E' sont dits *congrus* relativement à E'; nous noterons cette relation :

$$X \equiv Y \quad (\text{mod } E').$$

Nous avons évidemment

$$X \equiv X \quad (\text{mod } E')$$

et, d'autre part, les deux relations

$$X \equiv Y \quad (\text{mod } E'),$$
$$X \equiv Z \quad (\text{mod } E')$$

entraînent

$$X \equiv Z \quad (\text{mod } E').$$

Nous allons maintenant faire une abstraction et dire que tous les vecteurs congrus suivant E' définissent un seul vecteur $\mathfrak{X}$ d'un espace vectoriel $\mathcal{E}$ à $n - n'$ dimensions, appelé *espace projection de* E *parallèlement à* E'. D'une façon concrète nous pouvons associer à E' son sous-espace complémentaire E'', tel que

$$E = E' + E''.$$

Nous savons que dans ces conditions un vecteur quelconque X de E peut se décomposer en un vecteur de E' et un vecteur de E'', et cela d'une seule manière :

$$X = X' + X''.$$

Si $X \equiv Y \pmod{E'}$, puisque $X - Y = X' - Y' + X'' - Y''$, $X' - Y'$ étant un vecteur de E', $X'' - Y''$ étant un vecteur de E'', on aura nécessairement $X'' = Y''$. Si deux vecteurs X sont congrus suivant E', ils ont le même X''. Nous pouvons donc dire que X'' est une image de $\mathfrak{X}$ et que E'' est une image de $\mathfrak{E}$.

Revenons sur notre définition de $\mathfrak{E}$ et prouvons que l'ensemble des $\mathfrak{X}$ forme effectivement un espace vectoriel. En effet, la relation

$$X_1 \equiv X_2 \pmod{E'}$$

entraîne

$$aX_1 \equiv aX_2 \pmod{E'};$$

autrement dit, si $\mathfrak{X}$ appartient à $\mathfrak{E}$, il en est de même de $a\mathfrak{X}$. D'autre part, les relations

$$X_1 \equiv X_2 \pmod{E'}$$

et

$$Y_1 \equiv Y_2 \pmod{E'}$$

entraînent

$$X_1 + Y_1 \equiv X_2 + Y_2 \pmod{E'};$$

autrement dit, si $\mathfrak{X}$ et $\mathfrak{Y}$ appartiennent à $\mathfrak{E}$, il en est de même de leur somme $\mathfrak{X} + \mathfrak{Y}$.

Les opérations d'addition et de multiplication par un nombre s'appliquent aux vecteurs $\mathfrak{X}$ de $\mathfrak{E}$. Au vecteur $\mathfrak{X} = 0$ de $\mathfrak{E}$, correspond un vecteur quelconque de E'.

Cherchons le nombre de vecteurs linéairement indépendants de $\mathfrak{E}$, c'est-à-dire le nombre maximum h de vecteurs $\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_h$, non nuls, qu'il soit possible de trouver dans $\mathfrak{E}$, tels que

$$a_1\mathfrak{X}_1 + a_2\mathfrak{X}_2 + \ldots + a_h\mathfrak{X}_h = 0$$

entraîne la nullité de tous les a. Cela revient à chercher le nombre maximum h de vecteurs $X_1, X_2, \ldots, X_h$, n'appartenant pas à E', qu'il soit possible de trouver dans E, tels que la relation

$$a_1X_1 + a_2X_2 + \ldots + a_hX_h = X',$$

X' étant un vecteur quelconque de E', entraîne la nullité de tous les a. Or si à E' nous adjoignons E'', son sous-espace complémentaire dans E, tel que

$$E = E' + E'',$$

tout vecteur X_i de E se décompose d'une façon unique en la somme de deux vecteurs appartenant à chacun des deux sous-espaces :

$$X_i = X'_i + X''_i$$

et la relation devient

$$a_1 X''_1 + a_2 X''_2 + \ldots + a_h X''_h = 0.$$

Il en résulte que h est le nombre maximum de vecteurs linéairement indépendants de E''; donc $h = n'' = n - n'$.

Si E' a une dimension, E'' a $n - 1$ dimensions et constitue la projection ordinaire de E parallèlement à la direction de E'.

Nous dirons que, quel que soit a, l'ensemble des vecteurs aX constitue un même *rayon* qui sera défini par les rapports mutuels des $x_1, x_2, \ldots, x_n$ supposés non tous nuls.

9. Changement du système de coordonnées. — Soient $e_1, e_2, \ldots, e_n$, n vecteurs linéairement indépendants formant une base de E. Un vecteur X quelconque de E peut se mettre sous la forme

$$X = \sum_i x_i e_i. \tag{1}$$

Les nombres x_i sont des coordonnées relativement à ce système de base. Considérons maintenant n vecteurs $e'_1, e'_2, \ldots, e'_n$, de l'espace E, définis par

$$e'_k = \sum_{i=1}^{n} a_{ik} e_i.$$

Nous avons déjà vu que la condition nécessaire et suffisante pour qu'ils soient linéairement indépendants est que le déterminant $|a_{ik}|$ ne soit pas nul; ils forment alors un système de base, et l'on a

$$X = \sum_k x'_k e'_k. \tag{2}$$

Les x'_k sont les coordonnées du vecteur X par rapport au nouveau système de base, et nous avons

$$X = \sum_k x'_k \sum_i a_{ik} e_i = \sum_i e_i \sum_k a_{ik} x'_k,$$

soit, en identifiant avec la relation (1),

$$x_i = \sum_{k=1}^{n} a_{ik} x'_k.$$

Ce sont les formules de changement de coordonnées. Elles sont les

mêmes pour tous les vecteurs de l'espace E, dont les composantes se transforment d'une manière bien déterminée quand on fait le changement de base indiqué.

Toutes grandeurs dont la définition dépend du choix des axes de coordonnées, qui admettent des composantes et qui sont telles que ces composantes se transforment, dans le changement de base, à l'aide des formules ci-dessus, sont dites *covariantes entre elles*. Tous les vecteurs de E_n sont donc covariants.

10. Nous allons définir, en géométrie homogène linéaire, le *produit intérieur* de deux vecteurs X et Y [nous ne considérons pas cette expression comme synonyme de *produit scalaire*; nous réservons cette dernière expression pour une autre quantité introduite en géométrie métrique, à savoir $\sum_i \overline{x}_i y_i$]. Si X a pour coordonnées

$$x_1, x_2, \ldots, x_n \quad \text{et} \quad Y : y_1, y_2, \ldots, y_n,$$

le produit intérieur est, par définition,

$$x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

Nous avons les propriétés suivantes qui sont évidentes :

$$XY = YX,$$

$$x_i = Xe_i,$$

$$e_i e_k = \delta_{ik} \qquad (\text{c'est-à-dire o si } i \neq k, \text{ 1 si } i = k),$$

$$\alpha(XY) = (\alpha X)Y = X(\alpha Y) \qquad (\text{si } \alpha \text{ désigne un nombre}),$$

$$X(Y + Z) = XY + XZ,$$

et, plus généralement,

$$X\left(\sum_i Y_i\right) = \sum_i XY_i.$$

Considérons alors une forme linéaire :

$$f = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n;$$

elle peut être considérée comme le produit intérieur de deux vecteurs :

A de composantes $a_1, a_2, \ldots, a_n$;
X de composantes $x_1, x_2, \ldots, x_n$,

et s'écrire

$$f(X) = AX.$$

Si l'on a

$$X = \xi_1 X_1 + \xi_2 X_2 + \ldots + \xi_h X_h,$$

il s'ensuit que

$$f(X) = AX = A\left(\sum_i \xi_i X_i\right) = \sum_i \xi_i (AX_i) = \sum_i \xi_i f(X_i).$$

Si donc $f(X)$ est une forme linéaire, nous avons en particulier les deux propriétés :

$$f(aX) = af(X),$$
$$f(X + Y) = f(X) + f(Y).$$

Je dis que ces propriétés caractérisent les formes linéaires, c'est-à-dire que si à chaque vecteur X j'associe un nombre $L(X)$ tel que

$$(\mathrm{L}) \qquad \begin{cases} L(aX) = aL(Y), \\ L(X + Y) = L(X) + L(Y), \end{cases}$$

$L(X)$ est une forme linéaire.

En effet,

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$$

étant un vecteur quelconque, en appliquant successivement la seconde, puis la première propriété, il vient

$$L(X) = x_1 L(e_1) + x_2 L(e_2) + \ldots + x_n L(e_n),$$

où $L(e_1)$, $L(e_2)$, ..., $L(e_n)$ ne dépendent pas du vecteur X ; la propriété est bien établie. En résumé, une *forme linéaire* est une *fonction linéaire d'un vecteur* ou bien encore *un nombre attaché à un vecteur et satisfaisant aux relations fonctionnelles* (L).

III. — RÉSOLUTION ET DISCUSSION D'UN SYSTÈME D'ÉQUATIONS LINÉAIRES SANS DÉTERMINANTS [1].

1. Les éléments précédents de la théorie des vecteurs vont nous servir à exposer la méthode de Tœplitz pour résoudre et discuter un système linéaire de m équations à n inconnues.

---

(1) *Voir* Schmidt, *Rendiconti del Circolo Matematico di Palermo*, 1908; Tœplitz, *Ibid.*, 1909; Hasse, *Höhere Algebra* (Göschen, 1926).

Soit le système :

$$\text{N}(1)\quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = a_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = a_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = a_m. \end{cases}$$

Nous désignerons aussi le premier membre de la $i^{\text{ième}}$ équation par $f_i(X)$ ou $A_i X$.

Nous allons procéder à une première opération qui va « nettoyer » le système ou, d'une façon plus précise, le débarrasser des équations qui sont des combinaisons linéaires d'autres équations du système. A cet effet, nous cherchons le *nombre maximum de formes linéairement indépendantes* que l'on peut choisir parmi les m formes données aux premiers membres; soit r ce nombre, *rang* du système N(1) (certainement $r \leqq n$, $r \leqq m$); nous désignerons les formes correspondantes par $f_1, f_2, \ldots, f_r$. En employant un langage vectoriel, nous pouvons dire que nous cherchons le nombre maximum de vecteurs A linéairement indépendants et que nous désignons par $A_1, A_2, \ldots, A_r$ ceux d'entre eux qui peuvent être pris pour base. Un vecteur A autre que ceux-ci [ou une forme f autre que $f_1, f_2, \ldots, f_r$] est une combinaison linéaire des r premiers (*voir* § II, 6); on peut donc trouver des nombres λ tels que

$$\begin{aligned} f_{r+1} &\equiv \lambda_{11} f_1 + \ldots + \lambda_{1r} f_r, \\ &\ldots\ldots\ldots\ldots\ldots\ldots, \\ f_m &\equiv \lambda_{m-r,1} f_1 + \ldots + \lambda_{m-r,r} f_r. \end{aligned}$$

Si nous remplaçons x par une solution du système N(1), nous avons $f_i = a_i$ et par conséquent :

$$(\text{C})\quad \begin{cases} a_{r+1} = \lambda_{11} a_1 + \ldots + \lambda_{1r} a_r, \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ a_m = \lambda_{m-r,1} a_1 + \ldots + \lambda_{m-r,r} a_r. \end{cases}$$

Ce sont donc là des conditions *nécessaires* imposées aux seconds membres pour que le système N(1) présente des solutions; les solutions du système N(1), si elles existent, vérifient

$$\text{N}(2)\quad \begin{cases} f_1 = a_1, \\ \ldots\ldots, \\ f_r = a_r. \end{cases}$$

Montrons que, *réciproquement*, si les conditions C sont remplies, toute solution de N(2) est solution de N(1). Ceci est immédiat : si X

est solution de N(2), les r premières équations de N(1) sont satisfaites; d'autre part, considérons par exemple la $(r+1)^{\text{ième}}$ forme :

$$f_{r+1} \equiv \lambda_{11} f_1 + \ldots + \lambda_{1r} f_r;$$

quand on y remplace x par une solution de N(2), elle prend la valeur

$$\lambda_{11} a_1 + \ldots + \lambda_{1r} a_r,$$

qui est égale, d'après (C), à a_{r+1}; donc les $n-r$ dernières équations de (N_1) sont aussi satisfaites. Nous avons ramené notre système à ce qu'on appelle, dans la théorie avec déterminants, ses *équations principales :*

$$N(2) \quad \begin{cases} f_1 \equiv a_{11} x_1 + \ldots + a_{1n} x_n = a_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ f_r \equiv a_{r1} x_1 + \ldots + a_{rn} x_n = a_r. \end{cases}$$

2. Nous allons maintenant franchir une nouvelle étape. Définissons d'abord ce, qu'avec Tœplitz, nous entendons par *longueur d'une forme linéaire* à n variables; en supposant les n variables affectées des indices 1, 2, ..., n, nous désignerons par longueur d'une forme, la valeur de l'indice le moins élevé tel que toutes les variables affectées d'indices supérieurs aient dans la forme considérée un coefficient nul. La longueur suppose donc un ordre de numérotage imposé pour les variables. Une forme ne contenant que l'inconnue x_{p+1} est plus longue qu'une forme qui contiendrait effectivement $x_1, x_2, \ldots, x_p$, mais aucune autre variable.

Ceci étant posé, formons toutes les combinaisons possibles des r formes $f_1, f_2, \ldots, f_r$. Aucune n'est identiquement nulle puisque ces formes sont supposées linéairement indépendantes; leurs longueurs sont donc toutes comprises entre 1 et n; soit $k_1 (1 \leqq k_1 \leqq n)$ la longueur minima obtenue par (au moins) la forme

$$\gamma_1 = \gamma_{11} f_1 + \ldots + \gamma_{1r} f_r = \beta_{11} x_1 + \ldots + \beta_{1k_1} x_{k_1},$$

avec $\beta_{1k_1} \neq 0$ (sinon la forme serait de longueur moindre). Nous pouvons la *normer* en la multipliant par un facteur convenable, de façon à rendre le coefficient de x_{k_1} égal à l'unité :

$$g_1 = c_{11} f_1 + \ldots + c_{1r} f_r = b_{11} x_1 + b_{12} x_2 + \ldots + b_{1,k_1-1} x_{k_1-1} + x_{k_1}.$$

Il est évident qu'*il n'y a qu'une combinaison normale de longueur k_1*; car s'il y en avait deux, en les soustrayant on obtiendrait une combinaison linéaire de longueur moindre; ce qui est impossible

(puisque k_1 est la longueur minima), à moins d'obtenir un résultat identiquement nul, auquel cas la seconde combinaison serait identique à la première. Parmi les coefficients $c_{11}, c_{12}, \ldots, c_{1r}$, il y en a un au moins qui n'est pas nul; nous pouvons toujours supposer que c'est c_{1r} en interchangeant au besoin les indices des formes $f_1, f_2, \ldots, f_r$.

Nous allons maintenant recommencer les mêmes opérations sur les formes $f_1, f_2, \ldots, f_{r-1}$. Formons toutes leurs combinaisons linéaires; elles ont des longueurs comprises entre 1 et n dont le minimum est k_2; cette longueur correspond à une combinaison

$$\gamma_2 = \gamma_{21} f_1 + \gamma_{22} f_2 + \ldots + \gamma_{2,r-1} f_{r-1},$$

que nous normons :

$$g_2 = c_{21} f_1 + c_{22} f_2 + \ldots + c_{2,r-1} f_{r-1} = b_{21} x_1 + b_{22} x_2 + \ldots + b_{2,k_2-1} x_{k_2-1} + x_{k_2}.$$

Cette combinaison des $f_1, f_2, \ldots, f_{r-1}$ fait partie des combinaisons des $f_1, f_2, \ldots, f_r$. Par conséquent

$$k_2 \geqq k_1.$$

L'égalité est impossible, car il n'y a qu'une forme de longueur k_1 et les formes g_1 et g_2 ne peuvent être identiques en raison de la non-nullité de c_{1r} et de l'indépendance des formes $f_1, f_2, \ldots, f_r$; les formes g_1 et g_2 sont linéairement indépendantes : la dernière contient effectivement des variables que la première ne contient pas.

Et nous recommençons : parmi les coefficients $c_{21}, \ldots, c_{2,r-1}$, un au moins n'est pas nul, soit $c_{2,r-1}$. Nous écartons la forme f_{r-1} et nous formons les combinaisons linéaires de $f_1, f_2, \ldots, f_{r-2}$, ce qui nous conduit à un nombre $k_3 > k_2$, longueur d'une forme g_3, et ainsi de suite jusqu'à la $r^{\text{ième}}$ opération qui nous donne

$$g_r = c_{r1} f_1.$$

Parallèlement, formons les combinaisons des seconds membres

$$\begin{aligned} b_1 &= c_{11} a_1 + \ldots + c_{1r} a_r, \\ b_2 &= c_{21} a_1 + \ldots + c_{2,r-1} a_{r-1}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots, \\ b_r &= c_{r1} a_1. \end{aligned}$$

Les solutions de N(2) satisfont évidemment au *système récurrent*

$$\text{N(3)} \left\{ \begin{array}{ll} g_1 \equiv b_{11} x_1 + \ldots + b_{1,k_1-1} x_{k_1-1} + x_{k_1} & = b_1, \\ g_2 \equiv b_{21} x_1 + \ldots\ldots\ldots\ldots + b_{2,k_2-1} x_{k_2-1} + x_{k_2} & = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & , \\ g_r \equiv b_{r1} x_1 + \ldots\ldots\ldots\ldots\ldots\ldots + x_{k_r} & = b_r \end{array} \right.$$

avec

$$1 \leqq k_1 < k_2 < \ldots < k_r \leqq n.$$

Montrons que ce système N(3) est équivalent au système N(2), c'est-à-dire que toute solution de N(3) est aussi solution de N(2). En effet le système N(3) peut s'écrire

$$\begin{aligned}
g_1 &\equiv c_{11} f_1 + c_{12} f_2 + \ldots + c_{1r} \quad f_r \quad = c_{11} a_1 + c_{12} a_2 + \ldots + c_{1r} \quad a_r,\\
g_2 &\equiv c_{21} f_1 + c_{22} f_2 + \ldots + c_{2,r-1} f_{r-1} = c_{21} a_1 + c_{22} a_2 + \ldots + c_{2,r-1} a_{r-1},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\
g_r &\equiv c_{r1} f_1 \qquad\qquad\qquad\qquad\quad = c_{r1} a_1
\end{aligned}$$

Puisque $c_{r1} \neq 0$, la dernière relation donne $f_1 = a_1$ [première équation de N(2)]; $c_{r-1,2}$ n'étant pas nul, la précédente donne $f_2 = a_2$ [seconde équation de N(2)], et ainsi de suite jusqu'à la première qui donne $f_r = a_r$ [dernière équation de N(2)].

Pratiquement, pour obtenir le système N(3) de Tœplitz à partir du système N(2), nous procéderons de la façon suivante :

Cherchons la variable d'indice le plus élevé, qui n'est pas multipliée par un coefficient nul dans toutes les équations de N(2); ce sera x_{k_r} affectée d'un coefficient non nul dans une équation au moins; multiplions cette équation par un facteur convenable de façon à rendre égal à l'unité ce coefficient; nous obtiendrons

$$g_r = b_r.$$

De cette équation tirons x_{k_r} en fonction des variables d'indice moindre et reportons dans les $r-1$ autres équations; recommençons sur celles-ci les mêmes opérations jusqu'à obtenir une seule équation

$$g_1 = b_1.$$

Remarquons que de la relation

$$1 \leqq k_1 < k_2 < \ldots < k_r \leqq n,$$

nous déduisons $r \leqq n$, l'égalité n'étant possible que si les k sont formés de la suite des n premiers nombres entiers.

3. Nous sommes maintenant en mesure de discuter le système linéaire proposé N(1).

Si les conditions (C) ne sont pas remplies, nous avons déjà vu que N(1) est impossible; nous allons montrer que dans le cas contraire N(1) présente des solutions. En effet, alors, N(1) est équivalent à N(3)

et sur ce système la chose est évidente : choisissons arbitrairement x_1, x_2, $\ldots$, x_{k_1-1}, la première équation nous détermine x_{k_1}; choisissons encore arbitrairement x_{k_1+1}, x_{k_1+2}, $\ldots$, x_{k_2-1}, la deuxième équation nous détermine x_{k_2} et ainsi de suite. Nous avons ainsi obtenu une solution *particulière* du système proposé et nous savons que la solution *générale* est la somme de cette solution particulière et de la solution générale du système homogène (H_1)

$$(H_1)\quad \begin{cases} a_{11}\; x_1 + a_{12}\; x_2 + \ldots + a_{1n}\; x_n = 0, \\ a_{21}\; x_1 + a_{22}\; x_2 + \ldots + a_{2n}\; x_n = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}\, x_1 + a_{m2}\, x_2 + \ldots + a_{mn}\, x_n = 0. \end{cases}$$

La méthode de Tœplitz nous permet de démontrer que dans l'espace E_m à m dimensions, il ne peut y avoir $n > m$ vecteurs linéairement indépendants; car sinon, il y aurait m équations homogènes à $n > m$ inconnues qui ne devraient présenter que la solution o; or, d'après la forme de Tœplitz, il y a dans la résolution du système $n - r$ inconnues arbitraires avec $n - r > 0$, car $r \leqq m$ est inférieur à n.

Un système infini de vecteurs de E_n présente certainement une base, car le nombre de ses vecteurs linéairement indépendants est limité. Une infinité d'équations à n inconnues peut donc être ramenée à ses équations principales.

Les solutions du système (H_1) forment un système de vecteurs; parmi eux on peut en choisir au plus $s \leqq n$ linéairement indépendants, soient X_1, X_2, $\ldots$, X_s. Et une combinaison linéaire arbitraire X de ces s vecteurs est une solution du système homogène ([1]). L'ensemble des solutions du système homogène forme une multiplicité linéaire de base X_1, X_2, $\ldots$, X_s.

Nous allons montrer que $s = n - r$, n représentant le nombre de variables et r le rang du système. En effet, nous pouvons remplacer (H_1) par le système récurrent équivalent

$$(H_2)\quad \begin{cases} g_1 = 0, \\ g_2 = 0, \\ \ldots\ldots, \\ g_r = 0. \end{cases}$$

Si nous choisissons arbitrairement $x_1 = \xi_1$, $x_2 = \xi_2$, $\ldots$, $x_{k_1-1} = \xi_{k_1-1}$, la première équation détermine x_{k_1}; choisissons encore arbitrairement

([1]) On dit que X_1, X_2, $\ldots$, X_s forment un *système fondamental* de solutions de (H_1)

$x_{k_1+1} = \xi_{k_1}$, $x_{k_1+2} = \xi_{k_1+1}$, ..., $x_{k_2-1} = \xi_{k_2-2}$, la deuxième équation détermine x_{k_2}, etc. Par conséquent, on peut choisir arbitrairement toutes les variables autres que x_{k_1}, x_{k_2}, ..., x_{k_r}. Désignons ces autres variables par $x_{k_{r+1}}$, $x_{k_{r+2}}$, ..., x_{k_n}; la solution générale du système H_2 peut s'écrire

$$\begin{aligned}
x_{k_1} &= \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \ldots + \alpha_{1,n-r}\xi_{n-r},\\
x_{k_2} &= \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \ldots + \alpha_{2,n-r}\xi_{n-r},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\
x_{k_r} &= \alpha_{r1}\xi_1 + \alpha_{r2}\xi_2 + \ldots + \alpha_{r,n-r}\xi_{n-r},\\
x_{k_{r+1}} &= \xi_1,\\
&\ldots\ldots,\\
x_{k_n} &= \xi_{n-r},
\end{aligned}$$

les ξ étant des coefficients arbitraires. Nous avons donc $n-r$ *solutions fondamentales :*

$$\begin{array}{|c|c|c|c}
\alpha_{11} & \alpha_{12} & \ldots & \alpha_{1,n-r}\\
\alpha_{21} & \alpha_{22} & \ldots & \alpha_{2,n-r}\\
\ldots & \ldots & \ldots & \ldots\ldots\\
\alpha_{r1} & \alpha_{r2} & \ldots & \alpha_{r,n-r}\\
1 & 0 & \ldots & 0\\
0 & 1 & \ldots & 0\\
. & . & \ldots & .\\
0 & 0 & \ldots & 1
\end{array}$$

qui sont linéairement indépendantes, comme on le vérifie immédiatement par identification des $n-r$ dernières variables. Nous avons donc bien $s = n - r$.

Considérons le *système homogène transposé* (H'_1) :

$$(H'_1)\qquad \begin{cases}
a_{11}\lambda_1 + a_{21}\lambda_2 + \ldots + a_{m1}\lambda_m = 0,\\
a_{12}\lambda_1 + a_{22}\lambda_2 + \ldots + a_{m2}\lambda_m = 0,\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots;\\
a_{1n}\lambda_1 + a_{2n}\lambda_2 + \ldots + a_{mn}\lambda_m = 0.
\end{cases}$$

Soit $s' = m - r'$ le nombre de ses solutions linéairement indépendantes (r' étant le rang de H'_1); nous allons montrer que

$$s' = m - r,$$

c'est-à-dire que le *nombre de lignes linéairement indépendantes dans le tableau (ou matrice) des coefficients de* H_1 *est égal au nombre maximum de colonnes linéairement indépendantes*, ou encore que (H_1) et (H'_1) ont même rang :

$$r = m - s', \qquad r' = n - s \qquad \text{avec}\quad r = r'.$$

Désignons par r le nombre maximum des lignes linéairement indépendantes (rang de H_1) et affectons-les des premiers indices (*voir* au n° 1 de ce paragraphe)

$$\begin{array}{l} A_1(a_{11}, a_{12}, \ldots, a_{1n}), \\ A_2(a_{21}, a_{22}, \ldots, a_{2n}), \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ A_r(a_{r1}, a_{r2}, \ldots, a_{rn}). \end{array}$$

On a vu que

$$\begin{array}{ll} A_{r+1} = \lambda_{11} A_1 + \ldots + \lambda_{1r} A_r, & A_{r+1}(a_{r+1,1}, \ldots, a_{r+1,n}), \\ \ldots\ldots\ldots\ldots\ldots\ldots, & \ldots\ldots\ldots\ldots\ldots\ldots, \\ A_m = \lambda_{m-r,1} A_1 + \ldots + \lambda_{m-r,r} A_r, & A_m\ (a_{m,n}, \ldots, a_{m,n}). \end{array}$$

Le système homogène transposé (H'_1) présente donc les solutions suivantes :

$$\left|\begin{array}{l} \lambda_{1,1} \\ \lambda_{1,2} \\ \ldots \\ \lambda_{1,r} \\ -1 \\ 0 \\ . \\ 0 \end{array}\right| \begin{array}{l} \lambda_{2,1} \\ \lambda_{3,2} \\ \ldots \\ \lambda_{2,r} \\ 0 \\ -1 \\ \ldots \\ 0 \end{array} \quad \begin{array}{c} \ldots \\ \ldots \\ \ldots \\ \ldots \\ \ldots \\ \ldots \\ \ldots \\ \ldots \end{array} \left|\begin{array}{l} \lambda_{m-r,1} \\ \lambda_{m-r,2} \\ \ldots\ldots \\ \lambda_{m-r,r} \\ 0 \\ 0 \\ . \\ -1 \end{array}\right.$$

qui sont linéairement indépendantes (toujours en raison des $m - r$ dernières composantes); le nombre $m - r'$ de solutions linéairement indépendantes du système (H'_1) est donc supérieur ou égal à $m - r$, soit

$$m - r' \geqq m - r,$$

ou bien

$$r' \leqq r.$$

Mais en renversant les rôles du système homogène et du système homogène transposé, nous démontrerions de même que

$$r \leqq r'.$$

Donc

$$r = r'.$$

4. Nous pouvons donner aux conditions (C) (n° 1) une forme moins particulière; elles expriment que si X' est ***une des solutions fondamentales*** du système transposé, on doit avoir

$$X'A = 0,$$

où A est le vecteur $a_1, \ldots, a_m$ dont les composantes sont les deuxièmes membres de N(1).

Vu la linéarité de cette condition, il est équivalent de dire que A *doit être orthogonal aux solutions fondamentales du système transposé ou bien que* A *doit être orthogonal à toutes les solutions du système transposé.* Nous pouvons exprimer encore autrement cette dernière condition : pour que le système N(1) présente des solutions, il faut et il suffit que toute relation existant entre les premiers membres existe aussi entre les seconds membres.

Dans le cas particulier où $m = n$, toute cette étude nous conduit encore au *théorème de l'alternative :*

Si $r = n$, le système homogène n'a que zéro pour solution. Le système non homogène a toujours une solution et une seule.

Nous dirons, dans ce cas, que le tableau (ou la matrice) des coefficients est *régulier*.

Si $r < n$, le système homogène a des solutions non nulles, ainsi que le système homogène transposé. Le système non homogène n'est pas toujours possible; pour qu'il le soit, il faut et il suffit que les seconds membres forment un vecteur orthogonal à toutes les solutions du système transposé (ou, ce qui revient au même, aux solutions fondamentales de ce système); et si le système non homogène est possible, sa solution générale n'est pas unique puisqu'elle s'obtient en ajoutant à une solution particulière la solution générale de l'équation homogène. Nous disons ici que le tableau (ou matrice) du système est *irrégulier*, ou *singulier*, ou *dégénéré*.

CHAPITRE II.

I. — DÉVELOPPEMENTS DE GÉOMÉTRIE AFFINE. CALCUL DES MATRICES.

Nous avons vu dans le chapitre précédent l'effet du changement de base sur les coordonnées d'un vecteur de l'espace E_n. Soit e_1, e_2, ..., e_n l'ancienne base et soit ε_1, ε_2, ..., ε_n la nouvelle définie par

$$\varepsilon_k = a_{1k}e_1 + a_{2k}e_2 + \ldots + a_{nk}e_n \qquad \text{pour} \quad k = 1, 2, \ldots, n.$$

Si nous désignons par x_1, x_2, ..., x_n les anciennes coordonnées et par ξ_1, ξ_2, ..., ξ_n les nouvelles, nous avons la relation

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n = \xi_1 \varepsilon_1 + \xi_2 \varepsilon_2 + \ldots + \xi_n \varepsilon_n,$$

ou, d'une façon plus abrégée,

$$\sum_i x_i e_i = \sum_k \xi_k \varepsilon_k.$$

En remplaçant ε_k par son expression en fonction des e, il vient

$$\sum_i x_i e_i = \sum_k \xi_k \sum_i a_{ik} e_i = \sum_i e_i \sum_k a_{ik} \xi_k;$$

et puisque les e_i sont des vecteurs linéairement indépendants, cette égalité vectorielle est équivalente à

$$x_1 = \sum_k a_{1k} \xi_k,$$

$$x_2 = \sum_k a_{2k} \xi_k,$$

$$\ldots\ldots\ldots\ldots\ldots,$$

$$x_n = \sum_k a_{nk} \xi_k.$$

Les composantes x se déduisent donc des composantes ξ par une substitution linéaire A caractérisée par le tableau des coefficients ou *matrice*

$$\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{matrix}$$

et nous écrirons plus brièvement

$$A = \| a_{ik} \|.$$

Le déterminant $|a_{ik}|$ s'appelle le déterminant de A.

Remarquons que le déterminant $|a_{ik}|$ de cette matrice A n'est pas nul puisque les vecteurs ε_i sont linéairement indépendants; une telle matrice est dite *régulière*.

Remarquons encore que les *termes de la $k^{\text{ième}}$ colonne de la matrice sont les composantes du vecteur* ε_k ($k^{\text{ième}}$ vecteur coordonné nouveau) dans l'ancien système de coordonnées; cette interprétation d'une colonne de la matrice nous sera par la suite d'une grande utilité.

2. On peut donner à la matrice un autre sens plus général, puisqu'il permet d'envisager des matrices *irrégulières*; on considère les ξ comme étant les composantes d'un vecteur Ξ et les x comme étant les composantes d'un vecteur X *par rapport à un même système de coordonnées*; nous désignerons par A l'opération permettant de passer de Ξ à X et nous écrirons

$$X = A\,\Xi.$$

A s'appelle un *opérateur linéaire* [1] et définit une opération ou transformation linéaire pouvant s'appliquer à un vecteur quelconque de E_n; on s'assure immédiatement sur les formules développées données plus haut que cette transformation est bien linéaire, c'est-à-dire qu'au vecteur

$$\Xi = \Xi_1 + \Xi_2$$

elle fait correspondre le vecteur

$$X = X_1 + X_2,$$

---

[1] Nous emploierons en général indistinctement le mot opérateur ou matrice, bien qu'à vrai dire on doive distinguer entre l'opérateur *géométrique* et la matrice *algébrique* par laquelle s'exprime l'opérateur dans un certain système de coordonnées. Si le système de coordonnées change, l'opérateur géométrique restant invariant, la matrice change (*voir* § II).

X_1 étant le transformé de Ξ_1 et X_2 celui de Ξ_2 ; et qu'au vecteur

$$a\Xi$$

elle fait correspondre le vecteur

$$a X,$$

a désignant un nombre.

Un système de n^2 nombres a_{ik} définit un opérateur linéaire, *régulier* si le déterminant Δ n'est pas nul, *irrégulier* ou *singulier* ou *dégénéré* si le déterminant est nul. Les colonnes de la matrice ont encore une interprétation très simple : ce sont les *composantes des vecteurs transformés de* $e_1, e_2, \ldots, e_n$; par exemple le transformé ε_k de e_k a pour coordonnées

$$a_{1k}, \quad a_{2k}, \quad \ldots, \quad a_{nk}.$$

Il en résulte que, pour définir, analytiquement ou géométriquement, un opérateur A, il faut donner les n vecteurs ε_k transformés des vecteurs coordonnés par $A(\varepsilon_k = A e_k)$.

Les deux interprétations précédentes d'une matrice peuvent être *reliées l'une à l'autre* de la manière suivante :

Soient ξ_i les composantes de Ξ par rapport aux e_i; nous avons la relation

$$\Xi = \sum_i \xi_i e_i.$$

Appliquons la transformation *linéaire* A au vecteur Ξ; il vient

$$X = A\Xi = \Sigma \xi_i A e_i = \xi_i \varepsilon_i.$$

Par conséquent X a pour composantes ξ_i par rapport anx ε_i; en résumé, *si l'on opère sur les vecteurs de l'espace* E_n *et sur les vecteurs coordonnés la même transformation* A, *les coordonnées relatives des vecteurs ne sont pas changées par cette transformation.*

Dans la suite, il nous arrivera de *combiner les deux points de vue :* pour étudier un opérateur linéaire nous choisirons les vecteurs coordonnées les plus favorables, de façon que la matrice représentant l'opérateur soit la plus simple possible (représentations *canoniques*).

3. Égalité des matrices (ou opérateur). — Nous dirons que deux matrices (ou opérateurs) A et B sont égales lorsque, appliquées à un même vecteur *quelconque*, elles donnent le même résultat

$$AX = BX,$$

quel que soit X.

Les coordonnées correspondantes de ces deux vecteurs sont donc égales, soit

$$\sum_k a_{ik} x_k = \sum_k b_{ik} x_k \qquad \text{pour} \quad i = 1, 2, \ldots, n.$$

Ces relations, devant être vérifiées quelles que soient les valeurs attribuées aux x, entraînent

$$a_{ik} = b_{ik},$$

quels que soient i et k.

La condition nécessaire et suffisante pour que deux matrices soient égales est que les éléments correspondants soient égaux.

Somme de deux matrices (*ou opérateurs*). — Nous dirons qu'une matrice S est la somme de deux matrices A et B (et nous écrirons $S = A + B$), si à un vecteur quelconque X elle fait correspondre la somme des vecteurs transformés de X par A et par B

$$SX = AX + BX.$$

Soit

$$\sum_k s_{ik} x_k = \sum_k a_{ik} x_k + \sum_k b_{ik} x_k,$$

ce qui entraine

$$s_{ik} = a_{ik} + b_{ik}.$$

Par conséquent, pour faire la somme de deux matrices, on ajoute les éléments correspondants. La généralisation à la somme de plusieurs matrices est immédiate; cette opération est associative et commutative (ces deux propriétés étant vérifiées par la somme des éléments correspondants).

On définit immédiatement la *différence* de deux matrices A et B : c'est une matrice C qui ajoutée à B donne la matrice A,

$$A = B + C,$$

soit

$$a_{ik} = b_{ik} + c_{ik},$$

d'où

$$c_{ik} = a_{ik} - b_{ik}.$$

Par conséquent la soustraction de deux matrices est toujours possible et d'une seule manière; ses termes s'obtiennent en retranchant les termes correspondants des deux matrices données.

Nous dirons qu'une matrice est *nulle* si elle transforme un vecteur

quelconque en o, soit

$$\sum_k a_{ik} x_k = 0,$$

et puisque les x sont arbitraires : $a_{ik} = 0$. La condition nécessaire et suffisante pour qu'une matrice soit nulle est que tous ses éléments soient nuls.

Nous désignerons par *matrice unité*, et nous écrirons 1 ou E (Einheit, en langue allemande), une matrice qui appliquée à un vecteur quelconque ne le modifie pas. On voit immédiatement, par des considérations analogues aux précédentes, que la condition nécessaire et suffisante pour qu'une matrice soit égale à l'unité est que tous ses éléments soient nuls, sauf ceux situés dans la diagonale principale qui doivent être égaux à 1 ; plus simplement

$$a_{ik} = \delta_{ik}.$$

Produit de deux matrices (ou opérateurs). — Multiplier un vecteur par un nombre s, $(x_i = s\xi_i)$, c'est lui appliquer la *matrice scalaire*

$$s = \begin{Vmatrix} s & 0 & \dots & 0 \\ 0 & s & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & s \end{Vmatrix} = \| s\,\delta_{ik} \|.$$

La matrice A étant un opérateur linéaire, nous avons

$$s.AX = A(sX),$$

et ce vecteur est obtenu à partir du vecteur X par la matrice

$$sA = As = \| s a_{ik} \|,$$

que l'on appelle produit de la matrice A par le scalaire s. Évidemment $1.A = A.1 = A$.

Après avoir défini le produit d'une matrice par une matrice scalaire, nous allons définir le *produit de deux matrices quelconques*, soient

$$A = \| a_{ik} \|,$$
$$B = \| b_{ik} \|.$$

Si à un vecteur quelconque X, nous appliquons l'opérateur A, nous obtenons

$$Y = AX;$$

si à ce nouveau vecteur nous appliquons l'opérateur B, nous obtenons

$$Z = BY = BAX.$$

L'opération faisant passer de X à Z est linéaire; elle est obtenue par une matrice C, qui est dite *produit à gauche de la matrice* A *par la matrice* B et se représente par

$$C = BA \quad \text{(à lire de droite à gauche).}$$

Calculons l'expression de cette matrice

$$y_i = \sum_k a_{ik} x_k,$$

$$z_i = \sum_j b_{ij} y_j,$$

c'est-à-dire

$$z_i = \sum_j b_{ij} \sum_k a_{jk} x_k = \sum_k x_k \sum_j b_{ij} a_{jk};$$

ce qui nous conduit à

$$c_{ik} = \sum_j a_{ij} b_{jk}.$$

Règle. — *Pour multiplier deux matrices, on multiplie terme à terme les lignes de la première par les colonnes de la seconde afin d'obtenir un élément du produit.*

Il en résulte que, comme l'avait vu Cauchy,

$$(\text{dét } BA) = (\text{dét } B) \times (\text{dét } A);$$

le déterminant de la matrice produit est le produit des déterminants.

Le produit de deux matrices *n'est pas*, en général, une opération *commutative*. Considérons, par exemple,

$$A = \begin{Vmatrix} 1 & 1 \\ 0 & 1 \end{Vmatrix}, \qquad B = \begin{Vmatrix} 1 & 0 \\ 1 & 1 \end{Vmatrix},$$

nous aurons

$$C = BA = \begin{Vmatrix} 1 & 1 \\ 1 & 2 \end{Vmatrix},$$

tandis que

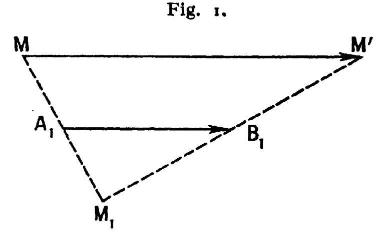
$$\Gamma = AB = \begin{Vmatrix} 2 & 1 \\ 1 & 1 \end{Vmatrix}.$$

Donnons un autre exemple de nature géométrique. La symétrie par rapport à un point est une opération linéaire représentable par une matrice; soit A la matrice représentant une symétrie par rapport au point A_1 et, de même, B la matrice représentant une symétrie par rapport au point B_1. La succession des deux symétries A et B est équi-

valente à la translation $2\overrightarrow{A_1B_1}$ et se représente par la matrice BA, tandis que la translation $2\overrightarrow{B_1A_1}$, différente de la précédente, se représente par AB.

Mais si nous considérons les symétries par rapport à deux droites rectangulaires, ce sont des opérations linéaires permutables; puisque leur succession équivaut à une symétrie par rapport au point de rencontre des deux droites. Le produit des deux matrices correspondantes est *commutatif*; nous dirons alors que ces deux matrices sont *permutables*. Si l'une des deux matrices est un nombre, autrement dit une matrice scalaire, il y a toujours permutabilité, ainsi que nous l'avons déjà montré.

Fig. 1.



Le produit de plusieurs matrices est une *opération associative*, c'est-à-dire que l'on a

$$C(BA) = (CB)A = CBA.$$

Cette propriété peut se vérifier par le calcul et l'on obtient comme résultat commun (1)

$$d_{ik} = \sum_{j,l} c_{ij} b_{jl} a_{lk},$$

mais la chose est évidente : considérons un vecteur X et appliquons-lui successivement les opérateurs A, B, C. Nous obtenons CBAX; mais il revient au même de lui appliquer A et B puis C, c'est-à-dire (BA),

(1) En effet

$$BA = \left\| \sum_l b_{jl} a_{lk} \right\| \quad \text{et} \quad C(BA) = \left\| \sum_j c_{ij} \sum_l b_{jl} a_{lk} \right\| = \left\| \sum_j \sum_l c_{ij} b_{jl} a_{lk} \right\|;$$

et de même

$$CB = \left\| \sum_j c_{ij} b_{jl} \right\| \quad \text{et} \quad (CB)A = \left\| \sum_l \left(\sum_j c_{ij} b_{jl} \right) a_{lk} \right\| = \left\| \sum_j \sum_l c_{ij} b_{jl} a_{lk} \right\|.$$

puis C et nous obtenons $C(BA)X$ ou bien A puis (CB); et nous obtiendrions $(CB)AX$. Nous avons donc, quel que soit X,

$$C(BA)X = (CB)AX = CBAX,$$

c'est-à-dire, par définition même, l'égalité des matrices $C(BA)$ et $(CB)A$.

Le produit est *distributif* par rapport à l'addition, c'est-à-dire que

$$(A + B)C = AC + BC$$

et

$$C(A + B) = CA + CB.$$

Ici encore ces résultats peuvent se vérifier par le calcul, mais ils sont presque évidents : au vecteur $Y = CX$ appliquons d'une part $A + B$ et d'autre part A et B; sommons ces deux vecteurs; d'après la définition même de la somme de deux matrices nous obtenons $(A + B)Y = AY + BY$, c'est-à-dire

$$(A + B)CX = ACX + BCX = (AC + BC)X,$$

pour un vecteur quelconque X, soit

$$(A + B)C = AC + BC.$$

Remarquons que

$$A \times 0 = 0 \times A = 0,$$
$$A + 0 = 0 + A = 0.$$

Et surtout notons bien le fait suivant : $AB = 0$ *n'entraîne pas nécessairement* $A = 0$ ou $B = 0$; par exemple, le produit de deux matrices

$$A = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}, \qquad B = \begin{Vmatrix} 0 & 0 \\ 0 & 1 \end{Vmatrix},$$

est $AB = 0$. Géométriquement : dans le plan, A représente la projection du vecteur sur le premier axe (parallèlement au deuxième axe), B sa projection sur le second (parallèlement au premier axe). La suite de ces deux opérations conduit à o quel que soit le vecteur d'où nous soyons partis.

Mais, *dans le cas où le déterminant de* A *n'est pas nul, la relation* $AB = 0$ *entraîne* $B = 0$. En effet on a

$$ABX = 0 \quad \text{quel que soit } X,$$

et, si nous posons $Y = BX$,

$$AY = 0,$$

qui, d'après l'hypothèse sur le déterminant de A, entraîne nécessaire-

ment $Y = o$; par conséquent BX est nul, quel que soit le vecteur X, autrement dit $B = o$.

Puissances d'une matrice. — Par définition, nous poserons

$$\begin{aligned} A^1 &= A \quad \text{et} \quad A^0 = 1 \quad \text{ou} \quad E, \\ A^2 &= A^1.A = AA, \\ A^3 &= A^2.A = AAA, \\ &\dots\dots\dots\dots\dots\dots, \\ A^m &= A^{m-1}.A = \underbrace{AA\ \dots\ A}_{m \text{ facteurs}}, \end{aligned}$$

Puisque nous savons définir une puissance positive quelconque de matrice, nous sommes en mesure de définir un polynome par rapport à une matrice. Soit le polynome

$$P(z) = \sum_{\nu=0}^{h} p_\nu z^\nu.$$

Nous pouvons calculer sans aucune ambiguïté

$$P(A) = \sum_{\nu=0}^{h} p_\nu A^\nu,$$

A désignant une matrice. Remarquons qu'il n'en serait pas de même pour un polynome à deux variables

$$P(z, u) = \sum_{\mu\nu} p_{\mu\nu} z^\mu u^\nu,$$

dans lequel nous substituerions aux deux variables, deux matrices; l'expression $P(A, B)$ ne serait bien définie que si les deux matrices A et B était permutables.

4. **Quelques opérateurs ou matrices particuliers.** — *a.* On appelle *matrice diagonale*, une matrice dans laquelle tous les éléments sont nuls, sauf ceux de la diagonale principale

$$A = \begin{Vmatrix} a_1 & o & \dots & o \\ o & a_2 & \dots & o \\ . & .. & \dots & . \\ o & o & \dots & a_n \end{Vmatrix} = \| a_{ik} = a_i \delta_{ik} \|.$$

Les matrices diagonales jouissent d'une propriété intéressante qui est

la suivante : ***deux matrices diagonales sont permutables et leur produit est une matrice diagonale***; vérifions-le :

$$A = \| a_i \delta_{ik} \|,$$
$$B = \| b_i \delta_{ik} \|,$$
$$AB = \left\| \sum_l a_i \delta_{il} b_l \delta_{lk} \right\| = \| a_i b_i \delta_{ik} \|.$$

Nous aurions d'ailleurs pu le voir encore plus simplement en revenant à la définition des matrices :

$$Y = BX, \qquad y_i = b_i x_i,$$
$$Z = AY, \qquad z_i = a_i y_i,$$
$$Z = ABX, \qquad z_i = a_i b_i x_i.$$

b. On appelle *matrice transposée* A′ d'une matrice A la matrice obtenue en permutant lignes et colonnes, soit

$$A = \| a_{ik} \|,$$
$$A' = \| a'_{ik} \| \quad \text{avec} \quad a'_{ik} = a_{ki}.$$

On appelle *matrice conjuguée* $\overline{A}$ d'une matrice A, la matrice obtenue en remplaçant chaque terme par son conjugué, soit

$$A = \| a_{ik} \|,$$
$$\overline{A} = \| \overline{a_{ik}} \|.$$

On appelle *matrice associée* A^* d'une matrice A, la conjuguée de la matrice transposée de A, soit

$$A = \| a_{ik} \|,$$
$$A^* = \| a^*_{ik} \|,$$

avec $a^*_{ik} = \overline{a_{ki}}$, $A^* = \overline{A'}$.

Il est évident que chacune de ces trois opérations appliquée deux fois successivement conduit à la matrice primitive :

$$(A')' = A,$$
$$\overline{\overline{A}} = A,$$
$$A^{**} = A.$$

Nous allons maintenant appliquer ces opérations à un produit de matrices ; nous avons

$$\overline{AB} = \overline{A} \cdot \overline{B}, \qquad \text{car} \quad \overline{\sum_j a_{ij} b_{jk}} = \sum_j \overline{a_{ij}}\, \overline{b_{jk}}.$$

La conjuguée d'un produit est donc le produit des conjuguées. Pour les deux autres opérations, le résultat est un peu moins simple.

Soient

$$A = \| a_{ik} \|, \quad \text{donc} \quad A' = \| \alpha_{ik} \| \quad \text{avec} \quad \alpha_{ik} = a_{ki},$$
$$B = \| b_{ik} \|, \quad \text{donc} \quad B' = \| \beta_{ik} \| \quad \text{avec} \quad \beta_{ik} = b_{ki};$$

et nous obtenons

$$AB = \left\| c_{ik} = \sum_l a_{il} b_{lk} \right\| \quad \text{et} \quad B'A' = \left\| \gamma_{ik} = \sum_l \beta_{il} \alpha_{lk} \right\|;$$

donc

$$\gamma_{ik} = \sum_l a_{kl} b_{li} = c_{ki},$$

et finalement

$$(AB)' = B'A'.$$

Par conséquent, pour calculer la matrice transposée d'un produit de matrices, il suffit de former les transposées de chaque facteur et d'en faire le produit en en renversant l'ordre. La même règle s'applique pour le calcul de la matrice associée; en effet prenons les conjugués des deux membres de la dernière relation obtenue

$$(AB)^* = \overline{B'A'} = \overline{B'}\,\overline{A'} = B^* A^*.$$

c. On appelle *matrice hermitienne* une matrice qui est égale à son associée

$$A = A^*, \quad \text{soit} \quad a_{ik} = \overline{a_{ki}}.$$

Si la matrice est réelle, c'est une matrice symétrique.

d. Nous verrons plus loin que les changements de coordonnées rectangulaires s'obtiennent par des matrices particulières que l'on appelle *unitaires*; une *matrice unitaire* U est caractérisée par la relation

$$UU^* = 1,$$

soit, si $U = \| u_{ik} \|$ et $U^* = \| u^*_{ik} = \overline{u_{ki}} \|$, la condition

$$\sum_j u_{ij} \overline{u_{kj}} = \delta_{ik}.$$

Nous démontrerons que deux matrices C et D satisfaisant à $CD = 1$ sont permutables; donc une matrice unitaire est encore caractérisée par la relation

$$U^* U = 1.$$

Par conséquent, *si la matrice* U *est unitaire, il en est de même de la matrice* U^* (rappelons que $U^{**} = U$).

Prenons les conjuguées des deux membres de la dernière relation, on a

$$\overline{UU^*} = 1$$

et finalement

$$\overline{U}\,\overline{U^*} = 1.$$

Donc, *si la matrice* U *est unitaire, il en est de même de la matrice* $\overline{U}$ $\left[\text{rappelons que } (\overline{U})^* = \overline{(U^*)}\right]$.

Si au lieu de prendre les conjuguées, nous avions pris les matrices transposées, nous aurions obtenu

$$U'(U^*)' = 1$$

ou

$$U'(U')^* = 1;$$

donc, *si* U *est unitaire, il en est de même de* U'.

5. Opérateur ou matrice inverse. — La matrice A, par la relation

$$Y = AX,$$

ou encore

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n, \end{aligned}$$

fait correspondre à un vecteur X de composantes $x_1, x_2, \ldots, x_n$, un vecteur et un seul Y de composantes $y_1, y_2, \ldots, y_n$. Si le déterminant de la matrice A, soit $|a_{ik}|$, n'est pas nul [1], on peut résoudre par rapport aux x_i ce système de n équations à n inconnues; et en désignant par A_{ik} le complément algébrique de a_{ik} dans le déterminant, les formules de Cramer nous donnent

$$x_i = \sum_k \frac{A_{ki}}{\Delta} y_k.$$

La matrice permettant de passer de Y à X se désigne par $A^{-1} = \|\alpha_{ik}\|$ et se nomme *matrice inverse* ou *réciproque* de A; l'on a

$$\alpha_{ik} = \frac{A_{ki}}{\Delta},$$

$$X = A^{-1}Y.$$

---

(1) Dét $A \neq 0$ signifie que les n vecteurs transformés de $e_1, e_2, \ldots, e_n$ par A (colonnes de A) sont linéairement indépendants.

Il est clair que si nous appliquons successivement la matrice A, puis la matrice A^{-1} à un vecteur X, nous obtenons un vecteur Y puis le vecteur X. La succession de ces deux opérations équivaut donc à une identité, ce que l'on écrit :

$$A^{-1}A = 1.$$

Le même raisonnement, où l'on permute l'ordre dans lequel on applique les deux matrices A et A^{-1}, nous conduit à la relation

$$AA^{-1} = 1.$$

La vérification par le calcul de ces identités est facile.

Le déterminant d'un produit de matrices est égal au produit des déterminants de ces matrices, car dans la règle du produit de deux matrices on reconnaît la règle du produit de deux déterminants; par conséquent,

$$\text{dét}(A) \times \text{dét}(A^{-1}) = 1.$$

6. Recherchons directement les *matrices réciproques*. Si la matrice A est singulière (si son déterminant est nul) elle ne peut avoir de *réciproque*, c'est-à-dire qu'il ne peut exister de matrice B telle que

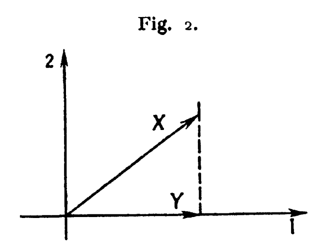
$$AB = 1,$$

sinon nous aurions la relation

$$\text{dét}(A) \times \text{dét}(B) = 1,$$

ce qui est impossible, puisque le premier facteur est nul.

Fig. 2.



Ces matrices singulières ou dégénérées jouent un rôle important, elles comprennent en particulier les *projections*. Exemple :

$$A = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}; \qquad Y = AX;$$

$$y_1 = x_1,$$

$$y_2 = 0.$$

Nous allons montrer qu'*à une matrice dégénérée* A, *il est possible d'associer une matrice* B *non dégénérée telle que*

$$A = P.B,$$

ou

$$A = B.P,$$

P *étant une matrice de projection.*

Soit $Y = AX$ qui symbolise l'ensemble des n relations suivantes :

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n. \end{aligned}$$

La matrice A étant supposée dégénérée, parmi ces n formes linéaires, r seulement sont linéairement indépendantes. Et par conséquent X décrivant E_n, Y appartient à un sous-espace E_r à r dimensions seulement, ce que nous écrirons

$$E_r = AE_n.$$

Nous allons montrer que l'on peut déterminer une matrice C telle que, *dans ce sous-espace*, $AC = 1$, ce qui veut dire que la transformation AC se réduit à la transformation identique lorsqu'elle est appliquée à un vecteur du sous-espace. Pour faire cette démonstration, effectuons un changement de coordonnées tel que les r nouveaux premiers vecteurs coordonnés e_1, e_2, ..., e_r, appartiennent au sous-espace E_r. De ce choix, il résulte que chaque équation :

$$\begin{aligned} AC_1 &= e_1, \\ AC_2 &= e_2, \\ &\ldots\ldots, \\ AC_r &= e_r \end{aligned}$$

pour déterminer respectivement les vecteurs C_1, C_2, ..., C_r est possible. D'autre part les équations

$$\begin{aligned} AC_{r+1} &= 0, \\ AC_{r+2} &= 0, \\ &\ldots\ldots, \\ AC_n &= 0 \end{aligned}$$

sont possibles, et les vecteurs C_{r+1}, ..., C_n peuvent être choisis linéairement indépendants, puisque le système homogène de n équations correspondant à l'équation $AC_i = 0$ étant de rang r, admet un système fondamental de $n - r$ solutions. C_{r+1}, C_{r+2}, ..., C_n forment donc une

base du système homogène, et la solution générale de chaque équation précédente en $C_1, C_2, \ldots, C_r$ s'obtient en ajoutant à une solution particulière la solution générale du système homogène. L'ensemble des vecteurs

$$C_1, C_2, \ldots, C_r, C_{r+1}, \ldots, C_n$$

est linéairement indépendant, car s'il existait une relation

$$\lambda_1 C_1 + \lambda_2 C_2 + \ldots + \lambda_r C_r + \lambda_{r+1} C_{r+1} + \ldots + \lambda_n C_n = 0,$$

elle entraînerait, en appliquant la matrice A,

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_r e_r = 0,$$

ce qui exigerait $\lambda_1 = \lambda_2 = \ldots = \lambda_r = 0$, les e_i étant linéairement indépendants; et puisque $C_{r+1}, \ldots, C_n$ sont indépendants, on aurait de plus :

$$\lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_n = 0.$$

Entre les C_i, une relation linéaire à coefficients non tous nuls est donc impossible; la matrice

$$C = \| C_1 \quad C_2 \quad \ldots \quad C_n \|,$$

dont les colonnes sont constituées par les composantes des vecteurs C_i, n'est pas dégénérée. Nous avons, par ailleurs,

$$AC = \| AC_1 \quad AC_2 \quad \ldots \quad AC_n \|,$$

en utilisant toujours cette propriété très importante : les colonnes d'une matrice sont les composantes des vecteurs transformés des vecteurs coordonnés par la matrice.

D'après les équations déterminant les C_i, on a donc

$$AC = \| e_1 \quad e_2 \quad \ldots \quad e_r \quad 0 \quad 0 \quad \ldots$$

ou, sous une forme plus développée,

$$AC = \begin{matrix} \\ r\left\{\begin{matrix} \\ \\ \\ \\ \end{matrix}\right. \\ \\ \\ \\ \end{matrix} \left\| \begin{matrix} \overbrace{1 & 0 & \ldots & 0}^{r} & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot & \ldots & \cdot \\ \underbrace{0 & 0 & \ldots & 0}_{r} & 0 & \ldots & 0 \end{matrix} \right\|.$$

On voit que l'opération AC n'altère aucun des vecteurs $e_1, \ldots, e_r$, et par suite respecte chaque vecteur de E_r.

C'est un opérateur de projection, puisque son application à un vecteur revient simplement à annuler les $n-r$ dernières coordonnées du vecteur. Nous avons donc

$$AC = P.$$

La matrice C n'étant pas dégénérée, présente une inverse

$$B = C^{-1}$$

qui n'est pas dégénérée et la dernière relation s'écrit

$$A = P.B,$$

ce que nous voulions établir. Nous l'avons fait en changeant les axes de coordonnées, mais nous verrons plus loin que ce changement transforme une matrice A en une matrice $S^{-1}AS$ et le résultat que nous venons d'obtenir subsiste ([1]).

7. Nous allons maintenant raisonner sur les matrices *sans utiliser les déterminants.*

Soit la relation

$$Y = AX.$$

Si elle est résoluble en X par

$$X = BY,$$

quel que soit Y, nous avons

$$Y = ABY,$$

soit

$$AB = 1.$$

On dit que la matrice A *n'est pas dégénérée* et que B est sa *réciproque à droite.* Dans ce cas, où il y a une réciproque à droite, il n'y en a qu'une; supposons en effet qu'il y en ait deux B_1 et B_2; nous aurions alors

$$AB_1Y = Y,$$
$$AB_2Y = Y,$$

soit, en soustrayant,

$$A(B_1 - B_2)Y = 0 \text{ pour tout } Y.$$

Or la matrice A n'étant pas dégénérée, d'après le théorème de l'alter-

([1]) Nous pouvons imaginer aussi qu'on a dans tout ce qui précède raisonné sur les *opérateurs géométriques* A, C, P, B, ... La conclusion, établie dans un système de coordonnées particulières qui simplifie les calculs, est générale.

native, cette relation entraîne

$$(B_1 - B_2)Y = 0$$

quel que soit Y, c'est-à-dire

$$B_1 = B_2.$$

Si B est la réciproque à droite de A, elle en est aussi la réciproque à gauche, c'est-à-dire que

$$AB = 1$$

entraîne

$$BA = 1.$$

Voici la démonstration de Tœplitz : partons de la relation

$$AB = 1.$$

multiplions à droite par A les deux membres :

$$ABA = A.$$

Soit

$$A(BA - 1) = 0$$

et, en ajoutant $AB = 1$, il vient

$$A(BA + B - 1) = 1.$$

La matrice $B + BA - 1$ serait donc une réciproque à droite de A; B étant la seule réciproque à droite, nous avons

$$BA = 1.$$

En résumé, étant donné une matrice, deux cas sont possibles :

Ou bien elle ne présente pas de matrice réciproque, auquel cas elle est dite dégénérée;

Ou bien elle en présente une qui est alors unique et réciproque à droite et à gauche.

Ces résultats ne s'appliquent pas aux matrices infinies et nous verrons qu'au lieu de ces deux cas, quatre peuvent se présenter :

Une réciproque unique à droite et à gauche;
Aucune, ni à droite, ni à gauche;
Aucune à droite, mais une infinité à gauche;
Aucune à gauche, mais une infinité à droite.

8. Les quatre matrices A, A′, $\overline{A}$, A* sont simultanément dégénérées ou

non. — Cela résulte immédiatement de la considération de leurs déterminants qui sont simultanément nuls ou non.

Si A présente une inverse, il en est donc de même des matrices A', $\bar{A}$, A^* et ces inverses sont en relations simples avec la première :

$$AA^{-1} = 1;$$

prenons les transposées des deux membres :

$$(A^{-1})'.A' = 1,$$

soit

$$(A')^{-1} = (A^{-1})'.$$

Et de la même façon nous prouverions que

$$(\bar{A})^{-1} = (\overline{A^{-1}}),$$
$$(A^*)^{-1} = (A^{-1})^*$$

Envisageons maintenant deux matrices A et B non dégénérées, leur produit n'est pas dégénéré, ainsi qu'il résulte immédiatement de la considération des déterminants correspondants, et présente donc une réciproque; je dis que

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Il faut prouver que

$$B^{-1}A^{-1}AB = 1,$$

ce qui est évident, car $A^{-1}A = 1$ et il reste $B^{-1}B = 1$, satisfaite par la définition même des matrices réciproques. On voit facilement que cette démonstration s'applique à un produit d'un nombre quelconque de facteurs, d'où la règle : *pour prendre la réciproque d'un produit de matrices, on prend la réciproque de chaque facteur et on en inverse l'ordre.*

Par conséquent

$$(A^n)^{-1} = (A^{-1})^n,$$

que, par définition, on écrit A^{-n}. Nous avons donc à l'heure actuelle défini les puissances entières positives et négatives des matrices; il est aisé de vérifier qu'elles satisfont aux règles de calcul ordinaires des exposants.

Enfin, remarquons que les matrices unitaires déjà définies sont les matrices dont l'associée est égale à la réciproque.

9. Matrices non carrées. — C'est une notion utile mais non pas indispensable, les matrices non carrées pouvant se ramener à des matrices carrées par adjonction de lignes ou de colonnes de zéros.

Une matrice carrée transforme linéairement un espace E_n, en général en un autre espace E_n; imaginons une transformation *linéaire* d'un espace E_m dans un espace E_n, qui, aux composantes $x_1, x_2, \ldots, x_m$, d'un vecteur X fasse correspondre les composantes $y_1, y_2, \ldots, y_n$ du vecteur Y transformé; les y sont des fonctions des x qui ne sont autres que des formes linéaires (à cause du caractère linéaire de la transformation), soit

$$y_i = \sum_{k=1}^{m} a_{ik} x_k \qquad \text{pour} \quad i = 1, 2, \ldots, n,$$

que l'on écrit symboliquement

$$Y = AX;$$

$A = \| a_{ik} \|$ est une matrice non carrée à n lignes et m colonnes.

De même que pour les matrices carrées nous dirons que :

Deux matrices sont *égales* si elles font correspondre à un vecteur quelconque le même transformé, ce qui exige que les éléments correspondants de ces deux matrices soient égaux, condition évidemment suffisante.

Le produit d'une matrice par un nombre est la matrice obtenue en multipliant chaque élément par ce nombre.

Une matrice S est dite la *somme* de deux matrices A et B, si elle transforme un vecteur quelconque en la somme de ses transformés par la matrice A et par la matrice B; on a la somme de deux matrices ayant le même nombre de lignes et de colonnes par

$$s_{ik} = a_{ik} + b_{ik}.$$

Définissons encore le *produit* de deux matrices :

Soient une matrice A transformant un vecteur X de E_m, en un vecteur Y de E_n :

$$Y = AX,$$

et une matrice B transformant un vecteur Y de E_n en un vecteur Z de E_p :

$$Z = BY.$$

La transformation permettant de passer de X de E_m à Z de E_p est linéaire et s'obtient par la matrice $C = BA$:

$$Z = CX.$$

Calculons les éléments de la matrice C en explicitant les relations

précédentes :

$$y_i=\sum_{k=1}^{m} a_{ik}x_k \qquad \text{pour} \quad i=1,\,2,\,\ldots,\,n,$$

$$z_l=\sum_{i=1}^{n} b_{li}y_i \qquad \text{pour} \quad l=1,\,2,\,\ldots,\,p,$$

soit

$$z_l=\sum_{i=1}^{n} b_{li}\sum_{k=1}^{m} a_{ik}x_k=\sum_{k=1}^{m}\left(\sum_{i=1}^{n} b_{li}\,a_{ik}\right)x_k,$$

d'où le terme général de la matrice C :

$$c_{lk}=\sum_{i=1}^{n} b_{li}a_{ik}.$$

Et l'on dit (d'une façon concise) que pour effectuer le produit BA, on multiplie chaque ligne de B par chaque colonne de A. Remarquons que le produit de deux matrices BA *n'a de sens que si le nombre des colonnes de* B *est égal au nombre de lignes de* A.

Ainsi que nous l'avons signalé au début de ce paragraphe, la notion de matrice non carrée n'est pas indispensable, mais elle est utile. En effet :

1° Nous avons montré déjà que la $i^{\text{ième}}$ colonne d'une matrice carrée est le vecteur transformé par cette matrice du $i^{\text{ième}}$ vecteur coordonné. Considérons un vecteur X de E_m comme une matrice à m lignes et une colonne, soit

$$X=\begin{Vmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{Vmatrix}$$

et de même un vecteur Y de E_n comme une matrice à n lignes et à une colonne, soit

$$Y=\begin{Vmatrix} y_1 \\ y_2 \\ \cdot\cdot \\ y_n \end{Vmatrix}.$$

La relation

$$Y=AX$$

se traduit par

$$y_i=\sum_{k=1}^{m} a_{ik}x_k \qquad \text{pour} \quad i=1,\,2,\,\ldots,\,n$$

et par conséquent peut être considérée comme une *relation entre matrices* non carrées, AX étant le produit des matrices A et X.

2° Considérons alors la matrice A qui fait passer d'un vecteur X de E_m à un vecteur Y de E_n :

$$Y = AX$$

et changeons de coordonnées, dans chacun de ces deux espaces, par les deux matrices S et T non dégénérées :

$$X = S\Xi,$$
$$Y = TH.$$

Par quelle matrice $\mathfrak{A}$ s'exprime l'opérateur A dans ce nouveau système de coordonnées?

Y = AX nous donne

$$TH = AS\Xi.$$

La matrice T n'étant pas dégénérée, la matrice T^{-1} existe; multiplions à gauche par T^{-1} les deux membres de la dernière relation, il vient :

$$H = T^{-1}AS\Xi.$$

Et par conséquent, la nouvelle matrice correspondant à l'opérateur A est

$$\mathfrak{A} = T^{-1}AS.$$

T^{-1} a n lignes et n colonnes, A a n lignes et m colonnes, S a m lignes et m colonnes et $\mathfrak{A}$ a bien n lignes et m colonnes.

10. Décomposition des opérateurs ou matrices. — Revenons aux matrices carrées A transformant linéairement l'espace vectoriel à n dimensions en lui-même et supposons que nous connaissions un sous-espace linéaire à E' à n' dimensions qui soit invariant par la matrice A, c'est-à-dire tel que si l'on choisit arbitrairement un vecteur X de E', son transformé Y = AX appartienne aussi à E' (ce que nous avons eu l'occasion d'exprimer de la façon suivante : AE' est identique à E' ou est contenu dans E'). Nous décomposerons alors E en deux sous-espaces dont l'un est E' et l'autre son complémentaire E'' à $n'' = n - n'$ dimensions; puis nous *adapterons* le système de coordonnées de E à cette décomposition, autrement dit nous choisirons les n' premiers vecteurs coordonnés dans E' et par suite les n'' autres dans E''. La matrice relative à ce nouveau système de coordonnées prend une forme plus simple, car les n' premières colonnes représentant les vecteurs transformés des n' premiers vecteurs coordonnés (de E') ont leurs n'' dernières lignes nulles

(les transformés appartenant également à E′).

$$A = \left\| \begin{array}{c:c} \overbrace{\qquad A_1 \qquad}^{n'} & \overbrace{\begin{matrix}\cdots\cdots\cdots\\ \cdots\cdots\cdots\\ \cdots\cdots\cdots\\ \cdots\cdots\cdots\end{matrix}}^{n''} \\ \hdashline \begin{matrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & 0 \end{matrix} & A_2 \end{array} \right\| .$$

Un vecteur de E′ se transforme en un vecteur de E′ par la matrice A_1 à n' lignes et n' colonnes qui est dite *induite* dans E′ par la matrice A. L'espace projection $\mathcal{E}$ de E suivant E′ peut être défini par les n'' dernières coordonnées; les vecteurs transformés par A des n'' derniers vecteurs coordonnés ont pour composantes les éléments des n'' dernières colonnes. Mais les projections sur $\mathcal{E}$ de ces transformés ont pour composantes les éléments des n'' dernières colonnes limitées aux n'' dernières lignes (colonnes de A_2). De telle sorte que l'on peut dire : A *induit* dans $\mathcal{E}$ une transformation linéaire définie par la matrice A_2 ; ou bien si $\mathcal{X}$ et $\mathcal{Y}$ sont les projections sur $\mathcal{E}$ de X et Y liés par $Y = AX$, ils sont liés par la relation

$$\mathcal{Y} = A_2 \mathcal{X}.$$

Précisons un peu ceci; un vecteur quelconque X est la somme d'un vecteur de E′ d'un vecteur de E″ :

$$X = X' + X''.$$

Appliquons l'opérateur A :

$$AX = AX' + AX'';$$

or AX' appartient à E′; donc la projection sur E″, parallèlement à E′, de AX est la même que celle de AX'', transformé d'un vecteur appartenant à E″, $(AX)'' = (AX'')''$; la définition de A_2 exprime en outre que $(AX'')'' = A_2 X''$; par conséquent, si nous décomposons de même le vecteur Y :

$$Y = Y' + Y'',$$

nous obtiendrons

$$Y'' = A_2 X''.$$

En général E″ *n'est pas un sous-espace invariant par l'opérateur* A; dans le cas particulier où il l'est, il est clair que la matrice A se simplifie

et prend la forme

$$A = \left\|\begin{array}{cccc:cccc} & & & & 0 & 0 & \ldots & 0 \\ & A_1 & & & 0 & 0 & \ldots & 0 \\ & & & & \cdot & \cdot & \ldots & \cdot \\ & & & & 0 & 0 & \ldots & 0 \\ \hdashline 0 & 0 & \ldots & 0 & & & & \\ 0 & 0 & \ldots & 0 & & A_2 & & \\ \cdot & \cdot & \ldots & \cdot & & & & \\ 0 & 0 & \ldots & 0 & & & & \end{array}\right\|.$$

On écrit quelquefois

$$A = A_1 \dot{+} A_2$$

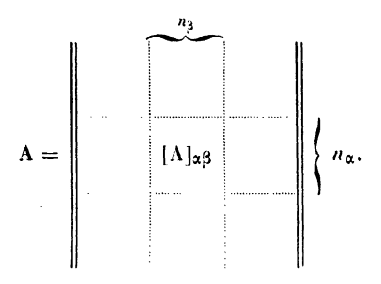
et l'on désigne cette opération par *somme directe*.

Ces résultats peuvent se généraliser. Supposons l'espace E, décomposé en E', E'', ..., sous-espaces *invariants* à n', n'', ... dimensions. Si nous adaptons le système de coordonnées à cette décomposition, la matrice se décompose en matrices carrées se succédant le long de la diagonale principale, ainsi qu'on s'en assure immédiatement en utilisant l'interprétation des colonnes d'une matrice, les éléments autres que ceux de A_1, A_2, A_3, ... étant nuls,

$$A = \left\|\begin{array}{ccc} A_1 & & O \\ & A_2 & \\ O & & A_3 \end{array}\right\|.$$

Ceci n'est plus vrai si les E', E'', ... ne sont pas invariants par A. Mais, de même que dans le cas simple de deux sous-espaces, nous pouvons envisager ce cas général où E est décomposé en sous-espaces E', E'', ..., $E^{(\alpha)}$, ..., $E^{(\beta)}$, ... et où le système de coordonnées est adapté à cette décomposition. Considérons un vecteur X de $E^{(\beta)}$ et son transformé $Y = AX$ par la matrice A; projetons-le sur $E^{(\alpha)}$ parallèlement à l'ensemble des sous-espaces autres que $E^{(\alpha)}$. Nous effectuons ainsi une transformation linéaire (à déterminant nul) qui nous conduit à $Y^{(\alpha)}$. En définitive, nous avons réalisé une transformation linéaire qui, à un vecteur X de $E^{(\beta)}$, fait correspondre un vecteur $Y^{(\alpha)}$ de $E^{(\alpha)}$; elle est représentée par une matrice $[A]_{\alpha\beta}$ à n_α lignes et n_β colonnes que nous

pouvons mettre en évidence dans la matrice A :



Les n_β vecteurs de base de $E^{(\beta)}$ se transforment par A en n_β vecteurs dont les coordonnées sont des colonnes de la matrice A; les coordonnées de leurs projections sur $E^{(\alpha)}$ s'obtiennent en ne conservant que les n_α lignes relatives aux composantes sur les vecteurs de base de $E^{(\alpha)}$ et en remplaçant les autres par des zéros. Finalement, la matrice $[A]_{\alpha\beta}$ à n_α lignes et n_β colonnes peut être prélevée dans la matrice A en ne gardant que les n_α lignes correspondant aux coordonnées relatives aux vecteurs de base de $E^{(\alpha)}$ et les n_β colonnes correspondant aux transformés des vecteurs de base de $E^{(\beta)}$ et en annulant tout le reste.

II. — TRANSFORMATION DES MATRICES. ÉQUIVALENCE DANS LE GROUPE LINÉAIRE HOMOGÈNE.

1. La transformation d'une matrice (ou opérateur) donnée A peut être envisagée de deux manières :

Ou bien, conservant toujours *le même opérateur géométrique*, nous changeons les axes de coordonnées, ce qui modifie la matrice correspondante;

Ou bien, nous soumettons l'espace dans lequel opère A à une transformation linéaire non dégénérée et nous cherchons l'opérateur transformé dans le nouvel espace.

Premier point de vue. — L'opérateur donné transforme un vecteur X en un vecteur Y :

$$Y = AX;$$

faisons le changement de coordonnées, défini par la matrice S (régulière) :

$$X = S\Xi,$$
$$Y = SH;$$

nous obtenons

$$SH = AS\Xi,$$

soit

$$H = S^{-1}AS\Xi.$$

Dans le nouveau système de coordonnées, l'opérateur s'exprime donc par la matrice :

$$\mathcal{A} = S^{-1}AS.$$

qu'on appelle *transformée de* A *par* S.

Nous verrons plus loin qu'on peut choisir S, c'est-à-dire les nouveaux vecteurs coordonnés de façon que $\mathcal{A}$ *soit plus simple que* A. C'est ce qui fait l'importance du point de vue actuel.

Les deux matrices A et $\mathcal{A}$ *correspondent au même opérateur géométrique sous des vêtements analytiques différents.* Si S est une matrice quelconque régulière, les deux matrices A et $S^{-1}AS$ sont dites *équivalentes dans le groupe linéaire homogène.*

Deuxième point de vue. — Effectuons sur l'espace E_n où opère A une transformation linéaire non dégénérée définie par la matrice S :

$$\mathcal{X} = SX,$$
$$\mathcal{Y} = SY.$$

La relation entre X et Y étant linéaire, celle qui relie $\mathcal{X}$ et $\mathcal{Y}$ le sera aussi. Et nous aurons la relation

$$\mathcal{Y} = \mathcal{A}\mathcal{X},$$

soit

$$SY = \mathcal{A}SX$$

ou

$$Y = S^{-1}\mathcal{A}SX = AX.$$

Donc

$$A = S^{-1}\mathcal{A}S.$$

Donc l'opérateur transformé de A dans le nouvel espace s'exprime par la matrice

$$\mathcal{A} = SAS^{-1}.$$

On peut dire que $\mathcal{A}$ est une *image de* A dans le nouvel espace.

2. Nous dirons que deux matrices A et B sont équivalentes, lorsqu'il est possible de trouver une matrice S, non dégénérée, telle que

$$B = S^{-1}AS.$$

L'ensemble des matrices ainsi transformées de la matrice A forme une

classe de matrices équivalentes dans le groupe homogène linéaire. Deux matrices de la même classe sont équivalentes; en effet

$$B = S^{-1}AS \quad \text{ou} \quad A = SBS^{-1},$$
$$C = T^{-1}AT,$$

qui peut s'écrire

$$C = T^{-1}SBS^{-1}T = (S^{-1}T)^{-1}B(S^{-1}T);$$

donc B et C sont équivalentes. De même, une matrice équivalente à une matrice de la classe appartient à la classe. La condition nécessaire et suffisante pour que deux matrices soient équivalentes est donc que les deux classes associées à chacune soient identiques.

Une classe est connue lorsqu'on en connaît une matrice; autrement dit à une matrice on peut faire correspondre une classe. Parmi toutes les matrices d'une classe, nous allons essayer d'en choisir une, *la plus simple possible*, de telle sorte qu'à une classe on puisse faire correspondre une matrice qui la caractérise. Cette matrice est équivalente à une matrice quelconque de la classe et en est appelée *matrice réduite* ou *canonique*. Il devient alors plus aisé de rechercher si deux matrices sont équivalentes :

La condition nécessaire et suffisante pour que deux matrices soient équivalentes est qu'elles aient même forme canonique ou réduite.

On peut dire encore que toutes les matrices d'une classe sont les différentes matrices par lesquelles s'exprime *un même opérateur* géométrique dans les divers systèmes de coordonnées.

Les coefficients de la matrice canonique sont des fonctions très importantes de la matrice A d'où l'on est parti; ces fonctions prennent même valeur pour toutes les matrices de la classe de A; nous mettons ainsi en évidence les *invariants de la matrice* A dans les transformations $S^{-1}AS$, S appartenant au groupe linéaire homogène.

Nous avons supposé que S est une matrice régulière arbitraire, et ceci nous a conduit à une *forme canonique dans le groupe des transformations de la géométrie homogène linéaire* et aux invariants d'une matrice dans ce même groupe de transformations. Mais on peut avoir à s'occuper de groupes de transformations plus restreints (par exemple, le groupe des rotations autour de l'origine), la matrice S est alors soumise à certaines conditions (dans cet exemple, relations entre les cosinus directeurs); la matrice canonique que l'on peut alors obtenir se différencie de la précédente.

Nous allons nous placer dans la géométrie homogène linéaire et par

conséquent prendre pour S la matrice régulière la plus générale; mais il était essentiel de remarquer que *la forme canonique et les invariants dépendent du groupe de transformations que l'on considère.*

3. Avant d'aborder le problème, démontrons quelques lemmes.

a. A ayant pour transformée $S^{-1}AS$, A^n a pour transformée

$$(S^{-1}A^nS = (S^{-1}AS)^n.$$

La vérification par le calcul est immédiate :

$$(S^{-1}AS)^n = \underbrace{S^{-1}AS.S^{-1}AS\ldots S^{-1}AS}_{n \text{ fois}}.$$

Tous les $SS^{-1} = 1$ intermédiaires disparaissent et il reste

$$(S^{-1}AS)^n = S^{-1}A^nS.$$

Géométriquement, la chose est évidente : A^n correspond à n fois la transformation A qui, après changement d'axes, donne n fois la transformation $S^{-1}AS$ ou bien $S^{-1}A^nS$.

Les transformés des puissances sont les puissances du transformé.

b. Si la matrice A a une réciproque A^{-1}, nous avons aussi

$$(S^{-1}AS)^{-1} = S^{-1}A^{-1}S$$

d'après la règle donnant la réciproque d'un produit.

Donc *prendre la réciproque d'une matrice et prendre sa transformée sont deux opérations permutables.*

c. Les deux systèmes homogènes :

$$AX = 0,$$
$$(S^{-1}AS)Y = 0$$

ont le même rang, c'est-à-dire le même nombre de solutions linéairement indépendantes. En effet, posons $Y = S^{-1}X$; la deuxième équation devient

$$S^{-1}AX = 0,$$

c'est-à-dire

$$AX = 0 \qquad (\text{car dét } S \neq 0)$$

et la correspondance linéaire biunivoque $Y = S^{-1}X$ entre les racines conserve l'indépendance linéaire.

d. Si $F(x)$ désigne un polynome, nous savons calculer cette expression lorsqu'on y remplace x par une matrice; sachant que la transformée d'une somme de matrices est la somme des transformées et que la transformée d'une puissance est la puissance de la transformée, nous en déduisons

$$F(S^{-1}AS) = S^{-1}.F(A).S.$$

e. Notons enfin que $\text{dét}(S^{-1}AS) = \text{dét}(SAS^{-1}) = \text{dét}A$; on retrouve le déterminant de A comme premier invariant.

4. Étant donnée une matrice A, cherchons s'il y a des vecteurs qui soient parallèles à leurs transformés par la matrice A (*directions invariantes*); ce qui revient à chercher un vecteur X tel que

$$AX = \lambda X,$$

soit

$$(\lambda - A)X = 0,$$

λ représentant la matrice :

$$\begin{Vmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & \lambda \end{Vmatrix}.$$

Pour que l'équation $(\lambda - A)X = 0$ ait des solutions X non nulles, il faut et il suffit que la matrice $\lambda - A$ soit dégénérée, c'est-à-dire que son déterminant soit nul; λ doit donc être racine de l'équation

$$f(\lambda) = 0,$$

$f(\lambda)$ étant le *polynome caractéristique* (¹) :

$$f(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & -a_{23} & \dots & -a_{2n} \\ -a_{31} & -a_{32} & \lambda - a_{33} & \dots & -a_{3n} \\ \dots\dots & \dots\dots & \dots\dots & \dots & \dots\dots \\ -a_{n1} & -a_{n2} & -a_{n3} & \dots & \lambda - a_{nn} \end{vmatrix} = \text{dét}(\lambda - A).$$

$$f(\lambda) = \lambda^n - s_1\lambda^{n-1} + s_2\lambda^{n-2} + \dots + (-1)^n s_n.$$

Si λ est un zéro de ce polynome, l'équation $(\lambda - A)X = 0$ présente une racine X non nulle au moins. Les racines de $f(\lambda) = 0$ s'appellent les

(¹) *Voir* dans CAUCHY, *Œuvres*, vol. 11, pages 75-133, l'introduction de cette notion pour l'étude des singularités des équations différentielles.

valeurs caractéristiques ou *valeurs propres* de A (les Allemands disent Eigenwert), et les solutions X correspondantes : *solutions propres* (Eigenlösung).

Montrons que $f(\lambda)$ *est invariant par un changement de coordonnées*. En effet, considérons

$$f(\lambda) = \text{dét}(\lambda - \mathfrak{A}),$$

soit

$$f(\lambda) = \text{dét}(\lambda - S^{-1}AS).$$

Remarquons que $\lambda = S^{-1}\lambda S$ (d'après le sens géométrique de la matrice scalaire λ); donc

$$f(\lambda) = \text{dét}(S^{-1}\lambda S - S^{-1}AS) = \text{dét}[S^{-1}(\lambda - A)S],$$

ou bien

$$f(\lambda) = \text{dét}(S^{-1})\,\text{dét}(\lambda - A)\,\text{dét}(S) = \text{dét}(\lambda - A).$$

Donc le polynome caractéristique de deux matrices équivalentes est le même; par conséquent les racines de l'équation caractéristique sont invariantes par rapport aux transformations linéaires et homogènes, ainsi que les expressions $s_1, s_2, \ldots, s_n$ qui sont des polynomes par rapport aux éléments de la matrice. En particulier :

s_n, déterminant de A, est un invariant.

Le plus important de ces invariants est

$$s_1 = a_{11} + a_{22} + \ldots + a_{nn} = \mathfrak{T}(A),$$

que l'on appelle *trace* (Spur) de la matrice A. Puisque la trace est une fonction linéaire des éléments, la trace de la somme de plusieurs matrices est la somme des traces de ces matrices. D'autre part, considérons une matrice $A = \| a_{ik} \|$ à n lignes et à m colonnes qui transforme l'espace E_m en un espace E_n et une matrice B qui transforme l'espace E_n en tout ou partie de l'espace E_m; B aura m lignes et n colonnes (certaines lignes pouvant n'être formées que de zéros). La matrice BA, à m lignes et m colonnes, transforme linéairement E_m en tout ou partie de lui-même. Et nous avons

$$\mathfrak{T}(AB) = \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} b_{ki},$$

$$\mathfrak{T}(BA) = \sum_{i=1}^{m} \sum_{k=1}^{n} b_{ik} a_{ki}.$$

Donc

$$\mathfrak{T}(AB) = \mathfrak{T}(BA).$$

5. Une théorie complète des invariants numériques de A par rapport au groupe homogène linéaire se ramène à celle des formes canoniques ou réduites de Jordan par rapport à ce même groupe ([1]) et dépasse notre programme; celle des invariants géométriques de l'espace lui est liée ([2]). Elle met en jeu la méthode des diviseurs élémentaires de Weierstrass; nous allons simplement en donner une idée.

Si l'équation $f(\lambda) = 0$ a ses n racines distinctes, à la racine λ_i correspond un seul vecteur X_i (à un scalaire près); nous prouverons que ces n vecteurs X_i sont linéairement indépendants et peuvent par conséquent être pris comme vecteurs coordonnés; puisque nous avons

$$AX_i = \lambda X_i,$$

la matrice se réduit alors à

$$\left\| \begin{matrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ . & \ldots & \ldots & . \\ 0 & 0 & \ldots & \lambda_n \end{matrix} \right\|.$$

Mais il est clair que les racines caractéristiques ne sont pas toujours distinctes; dans ce cas nous emploierons une méthode d'Hermann Weyl ([3]) développée par Schreier et qui conduit aux résultats de Jordan.

III. — NOTIONS SUR LES FORMES CANONIQUES OU RÉDUITES DE MATRICES DANS LE GROUPE LINÉAIRE HOMOGÈNE. RÉDUITE DE JORDAN.

1. Supposons que l'espace E se décompose en deux sous-espaces complémentaires E_1 et E_2, tels que chacun soit invariant par l'opérateur A. Si nous adaptons le système de coordonnées à cette décomposition de l'espace, nous avons vu que la matrice correspondante prend la forme

$$A = \left\| \begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_2 \end{array} \right\|,$$

([1]) *Voir* Jordan, *Traité des substitutions.*

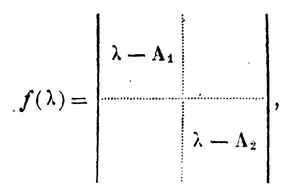
([2]) *Voir* Schreier et Sperner, *Vorlesungen über Matrizen*; Mac Duffee, *The Theory of matrices.*

([3]) H. Weyl, *Mat. Anal. der Raumproblems* (Springer, 1923).

que l'on écrit symboliquement

$$A = A_1 \dotplus A_2;$$

A_1 et A_2 s'appellent les opérateurs *induits par* A *dans les sous-espaces* E_1 et E_2. Il est clair que le polynome caractéristique prendra alors la forme



$f_1(\lambda) = \text{dét}(\lambda - A_1)$ représentant le polynome caractéristique de l'opérateur A_1 dans l'espace E_1, et $f_2(\lambda) = \text{dét}(\lambda - A_2)$ le polynome caractéristique de l'opérateur A_2 dans l'espace E_2. Développons ce déterminant par la règle de Laplace ([1]), relativement aux r premières colonnes, r désignant le nombre de dimensions de E_1; le seul mineur formé à l'aide de ces r premières colonnes, qui ne soit pas nul est,

$$\text{dét}(\lambda - A_1) = f_1(\lambda),$$

et son déterminant complémentaire est

$$\text{dét}(\lambda - A_2) = f_2(\lambda),$$

d'où la relation

$$f(\lambda) = f_1(\lambda) \times f_2(\lambda).$$

Nous avons donc montré que *si l'espace donné se décompose en deux sous-espaces invariants par* A, *le polynome caractéristique se décompose en un produit de polynomes, qui sont les polynomes caractéristiques des opérateurs induits par* A *dans les deux sous-espaces.*

2. Ce théorème présente une réciproque qui est très importante :

Si le polynome caractéristique relatif à un opérateur A *se décompose en deux facteurs premiers entre eux, dont les premiers coefficients sont égaux à l'unité, l'espace se décompose en deux sous-espaces disjoints* ([2]) *invariants par l'opérateur* A.

([1]) C'est une généralisation du développement classique par rapport aux éléments d'une colonne; on développe ici par rapport aux mineurs formés par p colonnes déterminées.

([2]) C'est-à-dire n'ayant en commun que le vecteur nul.

Nous allons d'abord établir quelques lemmes :

Premier lemme. — C'est seulement pour les racines $\lambda = \lambda_0$ de l'équation caractéristique $f(\lambda) = 0$, relative à A, qu'il existe des vecteurs X non nuls satisfaisant à la relation

$$AX = \lambda X \quad \text{ou} \quad (\lambda - A)X = 0.$$

Réciproquement, si $\lambda = \lambda_0$ est racine de $f(\lambda) = 0$, cette relation $(\lambda_0 - A)X = 0$ *définit une multiplicité linéaire de vecteurs* X.

Ce premier lemme est tout simplement la définition de l'équation caractéristique.

Deuxième lemme. — Étant donnée la matrice A, formons son équation caractéristique $f(\lambda)$, puis la matrice $f(A)$; nous allons montrer que *cette matrice appliquée à un vecteur arbitraire* X *le transforme en* 0, *autrement dit que* $f(A) = 0$. Ce théorème est *dû à Hamilton-Cayley.* Soit

$$f(\lambda) = \text{dét}(\lambda - A) = |\lambda\delta_{ki} - a_{ki}|,$$

et désignons par $b_{ik}(\lambda)$ le complément algébrique (¹); dans ce déterminant de $\lambda\delta_{ki} - a_{ki}$; $b_{ik}(\lambda)$ est manifestement un polynome de degré $n-1$ par rapport à λ et nous pouvons écrire

$$b_{ik}(\lambda) = \sum_{\nu=0}^{n-1} b_{ik}^{(\nu)} \lambda^\nu.$$

La matrice

$$B(\lambda) = \| b_{ik}(\lambda) \|$$

est appelée *matrice adjointe* de la matrice $\|\lambda - A\|$.

Nous avons

$$B(\lambda) = \left\| \sum_{\nu=0}^{n-1} b_{ik}^{(\nu)} \lambda^\nu \right\| = \sum_{\nu=0}^{n-1} \lambda^\nu \| b_{ik}^{(\nu)} \| = \sum_{\nu=0}^{n-1} B^{(\nu)} \lambda^\nu,$$

en posant

$$B^{(\nu)} = \| b_{ik}^{(\nu)} \|.$$

La relation

$$B(\lambda) = \sum_{\nu=0}^{n-1} B^{(\nu)} \lambda^\nu$$

exprime la matrice B sous forme d'un polynome en λ dont les coefficients

(¹) Nous disons indifféremment : *complément algébrique* ou *mineur affecté de son signe.*

sont des matrices. Formons alors le produit

$$B(\lambda)(\lambda - A) = \left\| \sum_l b_{il}(\lambda)(\lambda\delta_{lk} - a_{lk}) \right\|;$$

l'un des éléments de cette matrice,

$$\sum_l b_{il}(y)(\lambda\delta_{lk} - a_{lk}),$$

est le produit des éléments de la $k^{\text{ième}}$ colonne de dét$(\lambda - A)$ par les compléments algébriques des éléments de le $i^{\text{ième}}$ colonne. Il est donc nul si i est différent de k et égal à dét$(\lambda - A) = f(\lambda)$ si i est égal à k, soit

$$B(\lambda)(\lambda - A) = \| \delta_{ik} f(\lambda) \| = f(\lambda),$$

$f(\lambda)$ désignant une matrice scalaire. Nous avons ainsi obtenu une identité en λ :

$$\left(\sum_{\nu=0}^{n-1} B^{(\nu)} \lambda^\nu \right)(\lambda - A) = f(\lambda).$$

Les deux membres sont des polynomes de degré n en λ, à coefficients matriciels. Nous pouvons y remplacer λ par la matrice A; le premier membre s'annule et nous obtenons la relation cherchée

$$f(A) = 0.$$

Cette élégante démonstration est à rapprocher de celles employées ordinairement dans la théorie des déterminants adjoints.

La matrice A, opérant dans l'espace E_n, vérifie donc une équation algébrique de degré n; mais *il peut arriver que* A *vérifie une équation de degré moindre*. Parmi toutes ces équations il y en a une (si l'on impose au premier coefficient la condition d'être égal à l'unité) de degré minimum; son premier membre est appelé *polynome minimum de la matrice* A, et l'on démontre que *tout polynome qui s'annule lorsqu'on y remplace la variable par la matrice* A *est un multiple du polynome minimum*. En particulier, $f(\lambda)$ en est un multiple.

Nous sommes maintenant en mesure de démontrer la réciproque annoncée.

3. Théorème fondamental. — *Soit* A *une transformation linéaire de l'espace* E *et* $f(\lambda)$ *le polynome caractéristique correspondant; supposons qu'il se décompose en un produit de deux polynomes* $f_1(\lambda)$

et $f_2(\lambda)$, premiers entre eux, dont les premiers coefficients sont égaux à l'unité

$$f(\lambda) = f_1(\lambda) f_2(\lambda).$$

Dans ces conditions nous allons prouver qu'il existe deux sous-espaces E_1 *et* E_2 *jouissant des propriétés suivantes :*

1° E_1 *est invariant par* A; *il en est de même de* E_2.

2° E_1 *et* E_2 *sont disjoints, c'est-à-dire qu'ils n'ont que zéro comme vecteur commun.*

3° *Ils sont complémentaires :* $E = E_1 + E_2$.

4° f_1 *et* f_2 *sont les polynomes caractéristiques des transformations induites par* A *dans* E_1 *et* E_2.

5° *Enfin* E_1 *et* E_2 *sont déterminés d'une manière unique.*

D'après notre second lemme (Hamilton), l'espace E est défini comme formé par l'ensemble des vecteurs X satisfaisant à $f(A)X = 0$. Par conséquent, si $f_1(\lambda)$ est le polynome caractéristique de l'opérateur induit par A dans E_1, *nécessairement* E_1 sera défini par l'ensemble des vecteurs X tels que $f_1(A)X = 0$, et de même E_2 sera défini par l'ensemble des vecteurs X tels que $f_2(A)X = 0$. Ces relations définissent des espaces linéaires et la propriété (5) est bien vérifiée.

Reste à voir ce que contiennent les espaces E_1 et E_2 ainsi obtenus. A cet effet, envisageons les espaces $\mathcal{E}_1$ et $\mathcal{E}_2$ définis par

$$\mathcal{E}_1 = f_2(A)E, \quad \text{transformé de E par } f_2(A),$$
$$\mathcal{E}_2 = f_1(A)E, \quad \text{transformé de E par } f_1(A).$$

On a évidemment, d'après le théorème d'Hamilton-Cayley,

$$f_1(A)\mathcal{E}_1 = 0,$$
$$f_2(A)\mathcal{E}_2 = 0.$$

Donc E_1 contient $\mathcal{E}_1$ et E_2 contient $\mathcal{E}_2$. Nous allons montrer que E_1 est identique à $\mathcal{E}_1$ et E_2 à $\mathcal{E}_2$. Remarquons que $f_1(\lambda)$ et $f_2(\lambda)$ étant premiers entre eux, nous pouvons, d'après le théorème de Bezout, déterminer deux polynomes $g_1(\lambda)$ et $g_2(\lambda)$ tels que

$$f_1(\lambda)g_2(\lambda) + f_2(\lambda)g_1(\lambda) \equiv 1.$$

D'où, en remplaçant λ par la matrice A dans cette identité,

$$f_1(A)g_2(A) + f_2(A)g_1(A) = 1.$$

a. Supposons que $\mathcal{E}_1$ *soit vide*, autrement dit que $f_2(A)X$ soit nul,

quel que soit X. D'après la relation

$$X = f_1(A)g_2(A)X + f_2(A)g_1(A)X,$$

puisque dans l'expression $f_2(A)g_1(A)X$, nous pouvons permuter les facteurs polynomes qui ne dépendent que *d'un seul opérateur*, nous avons $f_2(A)g_1(A)X = 0$, donc

$$X = f_1(A)g_2(A)X = f_1(A)[g_2(A)X].$$

Par conséquent X appartient à $\mathcal{E}_2$; puisque $[g_2(A)X]$ est un vecteur de E; dans ces conditions $E \equiv \mathcal{E}_2$. La matrice $f_1(A)$ transforme donc E en lui-même, autrement dit $f_1(A)X = 0$ entraîne $X = 0$: E_1 est vide. Et d'autre part, $f_2(A)X$ étant nul quel que soit X, E_2 est identique à E.

En résumé, si $\mathcal{E}_1 = 0$, nous avons

$$\mathcal{E}_1 = E_1 = 0,$$
$$\mathcal{E}_2 = E_2 = E.$$

b. Supposons maintenant que ni $\mathcal{E}_1$ ni $\mathcal{E}_2$ ne soient vides; ces deux espaces sont disjoints en vertu de l'identité

$$X = g_2(A)f_1(A)X + g_1(A)f_2(A)X,$$

appliquée à un vecteur X commun, pour lequel

$$f_1(A)X = 0 \qquad (\text{car X est dans } \mathcal{E}_1),$$
$$f_2(A)X = 0 \qquad (\text{car X est dans } \mathcal{E}_2),$$

ce qui entraîne $X = 0$.

On a vu que E_1 contient $\mathcal{E}_1$ car d'après le théorème de Hamilton, $f_1(A)\mathcal{E}_1 = 0$. De même E_2 contient $\mathcal{E}_2$. Inversement, nous allons voir que $\mathcal{E}_1$ contient E_1; en effet, un vecteur X de E_1 satisfait à $f_1(A)X = 0$, donc à

$$X = f_2(A)g_1(A)X = f_2(A)[g_1(A)X]$$

et par conséquent appartient à $\mathcal{E}_1$ puisque $g_1(A)X$ est de E. Nous avons donc

$$\mathcal{E}_1 = E_1,$$
$$\mathcal{E}_2 = E_2,$$

mais, en introduisant $\mathcal{E}_1$ et $\mathcal{E}_2$, nous avons l'avantage d'une définition *explicite* des sous-espaces.

Un vecteur X quelconque se décompose

$$X = f_1(A)[g_2(A)X] + f_2(A)[g_1(A)X]$$

en la somme d'un vecteur de E_1 et d'un vecteur de E_2; E_1 et E_2 sont donc deux sous-espaces *complémentaires*.

E_1 et E_2 sont *invariants* relativement à A; soit en effet X un vecteur de E_1; il satisfait à

$$f_1(A)X = 0,$$

ou encore

$$Af_1(A)X = 0,$$

et en permutant l'ordre des facteurs, opération qui est légitime puisqu'il n'intervient qu'un seul opérateur A,

$$f_1(A)AX = 0,$$

ce qui nous montre que AX appartient à E_1.

Il nous reste à prouver que f_1 et f_2 sont les *polynomes caractéristiques* des transformations induites par A dans E_1 et E_2. Désignons en effet par φ_1 et φ_2 ces polynomes caractéristiques; d'après le théorème direct, nous avons

$$f(\lambda) = \varphi_1(\lambda)\varphi_2(\lambda).$$

Dans le cas particulier où E_1 serait vide, nous poserions $\varphi_1 = 1$. Nous allons montrer que f_1 et φ_1 ont les mêmes racines. Soit λ_0 une racine de $\varphi_1(\lambda) = 0$; nous avons

$$f_1(\lambda) = (\lambda - \lambda_0)\psi_1(\lambda) + f_1(\lambda_0)$$

en désignant par $\psi_1(\lambda)$ le quotient de $f_1(\lambda)$ par $\lambda - \lambda_0$. Remplaçons λ par la matrice A dans cette identité

$$f_1(A) = (A - \lambda_0)\psi_1(A) + f_1(\lambda_0).$$

λ_0 étant racine de $\varphi_1(\lambda) = 0$, il existe certainement dans E_1 un vecteur X_0 non nul tel que

$$AX_0 = \lambda_0 X_0;$$

nous avons alors

$$f_1(A)X_0 = \psi_1(A)(A - \lambda_0)X_0 + f_1(\lambda_0)X_0,$$

Or $f_1(A)X_0$ est nul, car X_0 appartient à E_1. Il en résulte que

$$0 = f_1(\lambda_0)X_0,$$

et X_0 n'étant pas nul, le scalaire $f_1(\lambda_0) = 0$. Toutes les racines de φ_1 appartiennent à f_1 premier avec f_2; donc φ_1 et f_2 sont premiers et, puisque

$$f_1 f_2 = \varphi_1 \varphi_2,$$

φ_1 divise f_1. Or de même φ_2 et f_1 sont premiers, donc f_1 divise φ_1. Finalement f_1 et φ_1 ne peuvent différer que par un facteur constant. Mais ces polynomes ayant tous deux 1 comme premier coefficient, nous avons

$$f_1 = \varphi_1,$$

et de même

$$f_2 = \varphi_2.$$

4. En appliquant le théorème fondamental de proche en proche, nous pouvons *l'étendre à un nombre quelconque de sous-espaces*. Soient μ_1, $\mu_2, \ldots, \mu_k$ les racines distinctes de $f(\lambda)$ et $n_1, n_2, \ldots, n_k$, leurs ordres de multiplicité

$$f(\lambda) = (\lambda - \mu_1)^{n_1}(\lambda - \mu_2)^{n_2} \ldots (\lambda - \mu_k)^{n_k}.$$

Si nous posons

$$f_1 = (\lambda - \mu_1)^{n_1} \qquad \text{et} \qquad f'_1 = (\lambda - \mu_2)^{n_2} \ldots (\lambda - \mu_k)^{n_k},$$

nous mettons en évidence le sous-espace E_1 défini par les vecteurs X tels que $(A - \mu_1)^{n_1} X = 0$ et un sous-espace E'_1 défini par $f'_1(A)X = 0$. Nous pouvons effectuer les mêmes opérations sur ce second sous-espace et ainsi de suite. Nous mettons finalement en évidence d'une manière *unique* k multiplicités linéaires $E_1, E_2, \ldots E_k$, définies respectivement par

$$(A - \mu_1)^{n_1} X = 0,$$
$$(A - \mu_2)^{n_2} X = 0,$$
$$\ldots\ldots\ldots\ldots\ldots,$$
$$(A - \mu_k)^{n_k} X = 0.$$

Ces multiplicités sont *non vides*, *disjointes*, *invariantes* par A, leur ensemble formant une décomposition de l'espace E, et $(\lambda - \mu_i)^{n_i}$ étant le *polynome caractéristique de la transformation* A_i *induite par* A *dans* E_i, ce qui nous prouve que E_i a n_i dimensions. Adaptons alors notre système de coordonnées à cette décomposition de l'espace en choisissant les vecteurs coordonnés

e_1,	e_2,	...,	e_{n_1}	dans E_1,
e_{n_1+1},	e_{n_1+2},	...,	$e_{n_1+n_2}$	dans E_2,
.....	.....,	...,	.....,	........,
e_{n-n_k+1},	e_{n-n_k+2},	...,	e_n	dans E_k.

La matrice A prend alors la forme

$$A = \left\| \begin{array}{cc:cc} \begin{array}{c:c} A_1 & 0 \\ \hdashline 0 & A_2 \end{array} & & 0 & \\ & & & \\ & 0 & \begin{array}{c:c} A_{k-1} & 0 \\ \hdashline 0 & A_k \end{array} & \end{array} \right\| \quad \begin{array}{l} n_1, n_2 \\ \\ n_{k-1}, n_k \end{array}$$

ce dont on s'assure immédiatement, une colonne de la matrice étant formée des composantes d'un vecteur transformé d'un vecteur coordonné par l'opérateur A. A_i est la matrice induite dans E_i par A et a pour polynome caractéristique $(\lambda - \mu)^{n_i}$. Dans le cas particulier où $f(\lambda)$ a toutes ses racines simples, la matrice A est ramenée à la forme diagonale

$$A = \left\| \begin{matrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 \\ \cdot & \cdot\cdot & \cdots & \cdots & \cdot\cdot \\ 0 & 0 & \dots & \lambda_{n-1} & 0 \\ 0 & 0 & \dots & 0 & \lambda_n \end{matrix} \right\|.$$

Dans le cas général, nous pouvons encore choisir arbitrairement les vecteurs coordonnés à l'intérieur de chaque sous-espace $E_1, E_2, \dots, E_k$. Nous allons profiter de cette latitude pour simplifier encore la matrice A et la ramener à la forme canoniqne de Jordan [1].

5. Nous nous proposons de simplifier la matrice A_1 de E_1 dont le polynome caractéristique est $(\lambda - \mu_1)^{n_1}$. Supprimons les indices pour abréger : dans un espace E à n dimensions nous avons un opérateur A qui, dans un certain système de coordonnées, s'exprime par une matrice A de polynome caractéristique $(\lambda - \mu)^n$; cherchons le système de coordonnées, par rapport auquel la matrice prenne la forme la plus simple possible.

La matrice $A_1 = A - \mu$ (ne pas confondre avec la notation A_1 de tout

---

(1) Ce problème est traité dans JORDAN, *Traité d'Analyse*, III, à l'occasion des points singuliers des équations différentielles. *Voir* aussi H. WEYL, *Math. Anal. des Raumproblems* (Springer, 1923).

à l'heure) a pour polynome caractéristique λ^n et est par conséquent dégénérée. D'après le théorème de Hamilton, nous avons la relation $A_1^n = 0$; nous sommes donc sûrs que les puissances de la matrice A_1 supérieures à la $n^{\text{ième}}$ sont nulles; désignons par *l le plus petit entier* (inférieur ou égal à n) *tel que* $A_1^l = 0$, et considérons les matrices non nulles :

$$A_1,\ A_1^2,\ \ldots,\ A_1^{l-1}.$$

Soient les multiplicités :

E_1 définie par les vecteurs X satisfaisant à $A_1^{l-1}X = 0$,
E_2 définie par ceux satisfaisant à $A_1^{l-2}X = 0$,
E_3 définie par ceux satisfaisant à $A_1^{l-3}X = 0$,
et ainsi de suite jusque E_{l-1} définie par ceux satisfaisant à $A_1 X = 0$.

Puisque $A_1^l = 0$, nous pouvons dire que $E_0 = E$ est définie par les vecteurs X satisfaisant à $A_1^l X = 0$.

Il est clair que *chacune de ces multiplicités est contenue dans celles qui la précèdent*, autrement dit que

$$E = E_0 > E_1 > E_2 > \ldots > E_{l-1} > 0,$$

car un vecteur de E_i satisfait à

$$A_1^{l-i}X = 0,$$

donc à

$$A_1^{l-i+1}X = 0$$

et par conséquent appartient à E_{i-1}. Enfin il est manifeste que E_{l-1} n'est pas vide, puisque définie par les vecteurs X satisfaisant à $A_1 X = 0$ avec une matrice A_1 dégénérée. Il est non moins évident que $A_1 E_i$ appartient à E_{i+1} car

$$A_1^{l-i}X = 0$$

entraîne

$$A_1^{l-(i+1)}(A_1 X) = 0.$$

Désignons par r_0 le nombre de dimensions de E_0 *relativement à* E_1, c'est-à-dire le nombre maximum de vecteurs de E_0 *linéairement indépendants* (module E_1) (¹); c'est, si l'on veut, le nombre de dimensions de la projection de E_0 parallèlement à E_1. Nous désignerons de même par r_1 le nombre de dimensions de E_1 relativement à E_2, par r_2 le

(¹) C'est-à-dire linéairement indépendants entre eux et des vecteurs définissant E_1; c'est-à-dire aussi tels que $\lambda_1 X_0^{(1)} + \ldots + \lambda_{r_0} X_0^{(r_0)} = X_1$ (X_1 de E_1) entraîne tous les λ_i nuls.

nombre de dimensions de E_2 relativement à E_3 et ainsi de suite jusqu'à r_{l-1} qui est le nombre de dimensions de E_{l-1} par rapport à zéro, c'est-à-dire tout simplement le nombre de dimensions de E_{l-1}.

Nous avons évidemment (Chap. I, § II, n° 8)

$$n = r_0 + n_1,$$

n_1 étant le nombre de dimensions de E_1; de même

$$n_1 = r_1 + n_2,$$
$$\ldots\ldots\ldots\ldots,$$
$$n_{l-1} = r_{l-1}.$$

Donc

$$r_0 + r_1 + \ldots + r_{l-1} = n.$$

D'autre part, E_1 défini par tous les vecteurs X satisfaisant à $A_1^{l-1} X = 0$, ne contient pas tout E, sans quoi nous aurions $A_1^{l-1} = 0$ et l ne serait pas l'entier minimum annulant A_1^l; donc $r_0 \geqq 1$. Nous allons montrer maintenant que la suite

$$r_0, r_1, r_2, \ldots, r_{l-1}$$

est non décroissante. Prenons dans E_0, r_0 vecteurs $X_0^{(1)}, X_0^{(2)}, \ldots, X_0^{(r_0)}$, linéairement indépendants (module E_1) ([1]). *Leurs transformés* $A_1 X_0^{(i)}$ *sont dans* E_1 (cette propriété a été établie ci-dessus) *et sont linéairement indépendants* (module E_2); supposons en effet que nous ayons la relation

$$\lambda_1 A_1 X_0^{(1)} + \ldots + \lambda_{r_0} A_1 X_0^{(r_0)} = X_2,$$

X_2 désignant un vecteur de E_2; appliquons la matrice A_1^{l-2} aux vecteurs des deux membres et nous obtenons

$$\lambda_1 A_1^{l-1} X_0^{(1)} + \ldots + \lambda_{r_0} A_1^{l-1} X_0^{(r_0)} = 0$$

ou encore

$$A_1^{l-1} [\lambda_1 X_0^{(1)} + \ldots + \lambda_{r_0} X_0^{(r_0)}] = 0,$$

ce qui signifie qu'il existe un vecteur X_1 de E_1 tel que

$$\lambda_1 X_0^{(1)} + \ldots + \lambda_{r_0} X_0^{(r_0)} = X_1,$$

relation qui entraîne la nullité des coefficients λ, les vecteurs $X_0^{(i)}$ ayant été choisis linéairement indépendants (module E_1). Nous venons d'établir

$$r_1 \geqq r_0.$$

([1]) Ils sont alors linéairement indépendants (mod E_k) car $E_k < E_1$. Donc en particulier

$$X_0^{(i)} \neq 0 \pmod{E_k}, \quad A_1^{l-k} X_0^{(i)} \neq 0 \qquad \text{pour} \quad k < l.$$

Aux $A_1 X_0^{(i)}$ adjoignons $s_1 = r_1 - r_0$ vecteurs de E_1, tels que les r_1 vecteurs ainsi formés soient linéairement indépendants (module E_2). Ces r_1 vecteurs forment les $X_1^{(j)}$ et satisfont donc à la relation

$$A_1^k X_1 \neq 0 \quad \text{si} \quad k < l - 1.$$

Leurs transformés $A_1 X_1^{(j)}$ sont dans E_2 et sont linéairement indépendants (module E_3) par la même démonstration que pour la multiplicité précédente; donc $r_2 \geqq r_1$.

Aux $A_1 X_1^{(j)}$ adjoignons $s_2 = r_2 - r_1$ vecteurs de E_2, tels que les r_2 vecteurs de E_2 ainsi formés soient linéairement indépendants (module E_3); ces r_2 vecteurs forment les X_2, etc.

Nous envisageons donc successivement les vecteurs :

Dans E_0, r_0 vecteurs X_0 :

$$X_0^{(1)}, \quad X_0^{(2)}, \quad \ldots, \quad X_0^{(r_0)}.$$

Dans E_1, r_1 vecteurs X_1 :

$$A_1 X_0^{(1)}, \quad A_1 X_0^{(2)}, \quad \ldots, \quad A_1 X_0^{(r_0)}, \quad X_1^{(r_0+1)}, \quad \ldots, \quad X_1^{(r_1)}.$$

Dans E_2, r_2 vecteurs X_2 :

$$A_1^2 X_0^{(1)}, \quad A_1^2 X_0^{(2)}, \quad \ldots, \quad A_1^2 X_0^{(r_0)}, \quad A_1 X_1^{(r_0+1)}, \quad \ldots, \quad A_1 X_1^{(r_1)}, \quad X_2^{(r_1+1)}, \quad \ldots, \quad X_2^{(r_2)}.$$

. . . .

Dans E_{l-1}, r_{l-1} vecteurs X_{l-1} :

$$A_1^{l-1} X_0^{(1)}, \quad A_1^{l-1} X_0^{(2)}, \quad \ldots, \quad A_1^{l-1} X_0^{(r_0)}, \quad A_1^{l-2} X_1^{(r_0+1)}, \quad \ldots,$$
$$A_1^{l-2} X_1^{(r_1)}, \quad A_1^{l-3} X_2^{(r_1+1)}, \quad \ldots, \quad A_1^{l-3} X_2^{(r_2)}, \quad \ldots, \quad X_{l-1}^{(r_{l-2}+1)}, \quad \ldots, \quad X_{l-1}^{(r_{l-1})}.$$

Les r_0 premiers (série X_0) sont linéairement indépendants (module E_1);
Les r_1 suivants (série X_1) sont linéairement indépendants (module E_2);
Les r_2 suivants (série X_2) sont linéairement indépendants (module E_3);
. . . .
Les r_{l-1} derniers (série X_{l-1}) sont linéairement indépendants entre eux.

Par conséquent, ces $r_0 + r_1 + \ldots + r_{l-1} = n$ vecteurs sont linéairement indépendants, car si l'on avait

$$\sum_{i=1}^{r_0} \lambda_i X_0^{(i)} + \sum_{j=1}^{r_1} \lambda_j X_1^{(j)} + \ldots + \sum_{k=1}^{r_{l-1}} \lambda_k X_{l-1}^{(k)} = 0,$$

le vecteur

$$\sum_{j=1}^{r_1} \lambda_j X_1^{(j)} + \ldots + \sum_{k=1}^{r_{l-1}} \lambda_k X_{l-1}^{(k)},$$

appartenant à E_1, nous aurions nécessairement $\lambda_i = 0$ pour $i = 1, 2, \ldots, r_0$ et la relation se réduirait à

$$\sum_{j=1}^{r_1} \lambda_j X_1^{(j)} + \ldots + \sum_{k=1}^{r_{l-1}} \lambda_k X_{l-1}^{(k)} = 0.$$

Le même raisonnement prouverait alors que $\lambda_j = 0$ et ainsi de suite. La proposition se trouve établie.

De plus les $A_1 X_0$ sont dans E_1, les $A_1 X_1$ dans $E_2, \ldots$, les $A_1 X_{l-1}$ sont nuls.

Nous allons maintenant ranger les vecteurs de ce tableau dans un autre ordre : par *colonnes*. Partons d'un vecteur X_0 et associons-lui sa *chaîne* d'images : $A_1 X_0$, $A_1^2 X_0$, $\ldots$, $A_1^{l-1} X_0$; nous pouvons ainsi former $s_0 = r_0$ chaînes comprenant chacune l vecteurs $\neq 0$ et commençant dans la multiplicité E_0. Puis les $s_1 = r_1 - r_0$ vecteurs X_1 qui ne sont pas de la forme $A_1 X_0$ donnent s_1 chaînes de $l - 1$ vecteurs commençant en E_1. Viendront ensuite $s_2 = r_2 - r_1$ chaînes de $l - 2$ vecteurs commençant en E_2, $\ldots$, et enfin $s_{l-1} = r_{l-1} - r_{l-2}$ chaînes de un vecteur de E_{l-1}.

Chacune des chaînes ainsi formées reste *invariante par* A_1, car chaque vecteur de la chaîne se transforme en celui qui le suit, et le dernier se transforme en zéro. Par conséquent, chaque chaîne sert de base à une multiplicité invariante par A_1; ces multiplicités sont au nombre de

$$s = s_0 + s_1 + \ldots + s_{l-1} = r_0 + (r_1 - r_0) + \ldots + (r_{l-1} - r_{l-2}) = r_{l-1}.$$

L'espace E est décomposé en r_{l-1} multiplicités disjointes et le système de coordonnées est adapté à cette décomposition. La matrice A_1 se décompose donc en s matrices carrées se succédant le long de la diagonale principale et qui comprendront : r_0 matrices à l lignes et l colonnes, $r_1 - r_0$ matrices à $l - 1$ lignes et $l - 1$ colonnes, $\ldots$, $r_{l-1} - r_{l-2}$ matrices à un élément. En utilisant le fait, déjà maintes fois signalé, que les éléments d'une colonne sont les composantes des vecteurs, transformés par la matrice, des vecteurs coordonnés, nous obtenons la forme de ces matrices partielles :

$$\begin{Vmatrix} 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{Vmatrix},$$

chaque vecteur coordonné étant transformé dans le suivant et le dernier en zéro.

Ceci pour la matrice A_1; quant à la matrice $A = A_1 + \mu$, elle se décompose en s matrices carrées se succédant le long de la diagonale principale; chacune du type suivant :

$$\begin{Vmatrix} \mu & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & \mu & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \mu & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ . & . & . & . & \dots & . & . \\ 0 & 0 & 0 & 0 & \dots & \mu & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & \mu \end{Vmatrix},$$

Il y a r_0 telles matrices à l lignes et l colonnes, $r_1 - r_0$ à $l - 1$ lignes et $l - 1$ colonnes, $r_2 - r_1$ à $l - 2$ lignes et $l - 2$ colonnes, ..., $r_{l-1} - r_{l-2}$ à un élément.

On désigne par *rang de* A_1 le nombre de dimensions de A_1E, c'est-à-dire le nombre de colonnes de la matrice A_1 linéairement indépendantes. Or il est évident que les colonnes non identiquement nulles sont indépendantes, et puisqu'il y a s colonnes de zéros (une par matrice partielle) le rang de A_1 est $n - s = n - r_{l-1}$.

Le nombre l, les nombres $r_0 \leqq r_1 \leqq r_2 \leqq \dots \leqq r_{l-1}$ tels que

$$r_0 + r_1 + \dots + r_{l-1} = n,$$

et le nombre μ, peuvent prendre toutes les valeurs; et il leur correspond chaque fois une matrice réduite de Jordan. Il suffit, en effet, d'écrire *a priori* la matrice qui correspond à ces nombres et de la transformer par une matrice S arbitraire non dégénérée; la matrice ainsi formée présente comme matrice réduite de Jordan la matrice désirée.

Deux matrices *équivalentes dans le groupe homogène linéaire* ont même matrice *réduite de Jordan* (et réciproquement), abstraction faite toutefois de l'ordre dans lequel se succèdent les matrices A_1, A_2,..., A_k.

6. Par conséquent les *invariants fondamentaux* d'une matrice A dans le groupe linéaire homogène sont : les invariants numériques μ_i (racines de l'équation caractéristique) et les invariants de structure : n_i (ordre de multiplicité des racines μ_i), l_i (donnant le nombre de lignes et de colonnes des matrices partielles de Jordan), $r_\nu^{(i)}$ (fixant le nombre de ces matrices partielles).

La condition nécessaire suffisante pour que deux matrices soient équi-

valentes dans le groupe linéaire homogène, est qu'elles aient les *mêmes invariants fondamentaux*.

Remarquons encore que si A est une matrice réelle et si les μ_i sont réels, la transformation de coordonnées donnant la forme réduite de Jordan est réelle et cette forme réduite est réelle.

7. Nous allons maintenant démontrer le théorème suivant ([1]) :

Pour qu'une matrice A *puisse se ramener à une matrice diagonale par un changement de coordonnées, il faut et il suffit que* $g(\mathrm{A}) = 0$, *en désignant par* $g(\lambda) = (\lambda - \mu_1)\ldots(\lambda - \mu_k)$ *le polynome ayant pour racines simples les racines distinctes de l'équation caractéristique relative à* A. g (A) *est alors le polynome minimum de* A. La réduction de A, réelle, à la forme diagonale, s'accomplit par un changement réel de coordonnées si, et seulement si, toutes les μ_i sont réelles.

Considérons en effet une matrice A équivalente à une matrice diagonale, et son polynome caractéristique

$$f(\lambda) = (\lambda - \mu_1)^{n_1}(\lambda - \mu_2)^{n_2}\ldots(\lambda - \mu_k)^{n_k};$$

nous avons vu que l'espace se décompose en k sous-espaces disjoints E_1, E_2, ..., E_k. Montrons que, *dans ce cas particulier*, E_i peut être défini par les vecteurs X_i satisfaisant à

$$(\mathrm{A} - \mu_i)\mathrm{X}_i = 0.$$

En effet, formons, pour fixer les idées, la matrice réduite équivalente à

$$\mathrm{A}_1 = \mathrm{A} - \mu_1 = \left\| \begin{array}{ccc|ccc} \mu_1 - \mu_1 & & & & & 0 \\ & \ddots & & & & \\ & & \mu_1 - \mu_1 & & & \\ \hline & & & & \cdots & \\ & & & & \mu_2 - \mu_1 & \\ & & & & \cdots & \\ 0 & & & & & \end{array} \right\|$$

nous vérifions bien que tous les vecteurs de E_1 (dont les n_1 premières coordonnées seules ne sont pas nulles) sont transformés en zéro par cette matrice, puisque les n_1 premières colonnes de celle-ci ne sont formées que de zéros. Il est clair que, réciproquement, tout vecteur satisfaisant à $(\mathrm{A} - \mu_i)\mathrm{X}_i = 0$ est bien un vecteur de E_i car il satisfait *a fortiori*

([1]) *Voir* SCHREIER et SPERNER, *Vorlesungen über Matrizen.*

à $(A - \mu_i)^{n_i} X_i = 0$. Or un vecteur quelconque X appartient à l'un des sous-espaces $E_1, E_2, \ldots, E_k$ et par conséquent vérifie la relation

$$(A - \mu_1)(A - \mu_2)\ldots(A - \mu_k)X = 0,$$

ce qui revient à dire que $g(A) = 0$.

Réciproquement, supposons cette relation satisfaite; un vecteur X_1 de E_1 satisfera à la relation

$$(A - \mu_1)X_1 = 0,$$

car, sinon, $Y_1 = (A - \mu_1)X_1$ serait un vecteur de E_1 non nul et l'on devrait avoir

$$(A - \mu_2)\ldots(A - \mu_k)Y_1 = 0,$$

relation impossible comme nous nous en assurons de proche en proche : $(A - \mu_2)Y_1 \neq 0$, ou bien Y_1 appartiendrait à E_2 qui par hypothèse est disjoint de E_1; donc $(A - \mu_2)Y_1 = (A - \mu_1)(A - \mu_2)X_1$ est un vecteur non nul et appartenant à E_1 et nous continuons. Tout vecteur de E_i satisfait à $(A - \mu_i)X_i = 0$, c'est-à-dire que la matrice induite par $A - \mu_i$ dans la multiplicité E_i est nulle, ce qui entraîne immédiatement $l_i = 1$ et par conséquent $r_0 = n_i$. La forme de Jordan est alors une forme diagonale, car elle ne comprend plus de sous-matrices d'ordre supérieur à 1, et le théorème est établi.

8. Donnons en terminant quelques indications sur les *matrices dégénérées*. Le déterminant étant nul, elle admet zéro pour valeur caractéristique, à l'ordre n_0.

Une telle matrice a un polynome caractéristique :

$$f(\lambda) = \lambda^{n_0}(\lambda - \mu_1)^{n_1}\ldots(\lambda - \mu_p)^{n_p}.$$

L'espace E se décompose en deux sous-espaces invariants : E_0 d'une part, $E' = E_1 + E_2 + \ldots + E_p$ d'autre part. Désignons par A_0 la matrice induite par A dans le premier sous-espace et par A' la matrice induite dans le second sous-espace; A_0 a pour équation caractéristique $\lambda^{n_0} = 0$ et est dégénérée, tandis que A' a pour équation caractéristique

$$(\lambda - \mu_1)^{n_1}\ldots(\lambda - \mu_p)^{n_p} = 0,$$

et n'est pas dégénérée.

Adaptons notre système de coordonnées à cette décomposition de l'espace en choisissant les premiers vecteurs coordonnés $e_1, e_2, \ldots, e_{n_0}$

dans E_0 et les autres $e_{n_0+1}, e_{n_0+2}, \ldots, e_n$ dans E′. La matrice **A** s'écrit alors

$$A = \left\| \begin{array}{c:c} A_0 & 0 \\ \hdashline 0 & A' \end{array} \right\|$$

et son carré

$$A^2 = \left\| \begin{array}{c:c} A_0^2 & \\ \hdashline & A'^2 \end{array} \right\|$$

et d'une façon générale sa puissance $k^{\text{ième}}$:

$$A^k = \left\| \begin{array}{c:c} A_0^k & \\ \hdashline & A'^k \end{array} \right\|$$

Le théorème de Hamilton-Cayley nous apprend que $A_0^{n_0} = 0$; désignons alors par $l \leqq n_0$ le plus petit entier positif tel que $A_0^l = 0$. La matrice A^l transforme donc E_0 en o et transforme E′ en lui-même biunivoquement puisque la matrice A'^l n'est pas dégénérée. Nous pouvons encore dire que *la matrice* A^l *transforme l'espace* E *en l'espace* E′; il en est évidemment de même des puissances supérieures à la $l^{\text{ième}}$ de la matrice A. Dans cet espace E′, qui est le *noyau* de E par rapport à la transformation A, *la transformation* A *est biunivoque et se réduit à* A′.

Remarquons que si $k < l$, la matrice A_0^k n'est pas nulle et par conséquent l'un des vecteurs : $A_0^k e_1, A_0^k e_2, \ldots, A_0^k e_{n_0}$, au moins, appartenant à E_0, n'est pas nul; donc

$$A^k E > E' \quad \text{si} \quad k < l,$$

tandis que

$$A^m E = E' \quad \text{si} \quad m \geqq l.$$

Il est d'autre part clair que chacun des sous-espaces $E_0, A_0E_0, A_0^2E_0, \ldots,$ est contenu dans le précédent; de telle sorte que

$$E_0 > A_0E_0 > A_0^2 E_0 > \ldots > A_0^{l-1} E_0 > A_0^l E_0 = 0;$$

ce qui donne, en remarquant que $AE = A_0E_0 + A'E'$,

$$E > AE > A^2E > \ldots > A^{l-1}E > A^lE = A^{l+1}E = A^{l+2}E = \ldots = E'.$$

Le noyau E′ est donc atteint, à partir de E, après $l \leqq n_0$ applications de la matrice A.

9. Remarquons encore une fois que la recherche de la forme réduite et des invariants dépend du groupe de transformations dans lequel on opère; dans l'étude précédente nous nous sommes placés dans le groupe homogène linéaire. Mais, plus loin, nous introduirons une métrique et nous nous limiterons au groupe unitaire (c'est-à-dire au groupe des transformations conservant les longueurs); ceci nous introduira de nouveaux invariants, mais surtout donnera un *sens géométrique* intéressant *aux racines de l'équation caractéristique* dont l'ensemble forme le spectre. L'introduction de la métrique nous permettra de poursuivre assez loin l'étude du spectre, mais le groupe unitaire nous conduira pour une matrice quelconque à une forme réduite beaucoup moins simple que la forme de Jordan.

IV. — NOTIONS SUR LE SPECTRE.

1. **Spectre.** — Nous désignerons sous le nom de *spectre* l'ensemble des *valeurs caractéristiques* λ_i, c'est-à-dire l'ensemble des $\lambda = \lambda_i$ satisfaisant à l'équation caractéristique

$$f(\lambda) \equiv \text{dét}(\lambda - A) = 0.$$

Il résulte de la façon même dont nous avons trouvé les valeurs caractéristiques que, si λ n'est pas égal à une valeur caractéristique, l'équation

$$\lambda X = AX$$

n'a pour solution que $X = 0$; tandis que si λ est égal à une valeur caractéristique, l'équation précédente présente au moins une solution non nulle, et dans le cas général elle en présente une infinité formant une multiplicité linéaire $\mathcal{E}_i = E^{(i)}_{l_i-1}$ ayant un nombre de dimensions $\nu_i = r_{l_i-1}$ supérieur ou égal à 1, inférieur ou égal à n_i. Ces vecteurs X s'appellent *solutions caractéristiques* ou *solutions propres*.

1° Il est clair que la condition nécessaire et suffisante pour que la matrice A soit dégénérée est que 0 soit une valeur caractéristique.

2° Supposons qu'il n'en soit pas ainsi. La matrice A présente alors une matrice inverse A^{-1} dont nous allons déterminer le spectre; λ_i étant une valeur caractéristique de A et X_i une solution caractéristique correspondante, nous avons

$$\lambda_i X_i = AX_i,$$

relation que nous pouvons multiplier à gauche par A^{-1} :

$$A^{-1}\lambda_i X_i = X_i,$$

soit encore

$$A^{-1}X_i = \frac{1}{\lambda_i} X_i.$$

Nous voyons donc que les inverses des valeurs caractéristiques de A sont valeurs caractéristiques de A^{-1}. La réciproque est évidente : il suffit de partir d'une valeur caractéristique μ_j de A^{-1}, $\frac{1}{\mu_j}$ sera valeur caractéristique de A. Nous avons donc établi que *les valeurs caractéristiques de* A^{-1} *sont les inverses des valeurs caractéristiques de* A, tandis que *les solutions caractéristiques correspondantes sont les mêmes.*

3° Effectuons, maintenant, un changement de coordonnées défini par $X = S\Xi$; l'opérateur A sera alors représenté par la matrice $\mathfrak{A} = S^{-1}AS$. La nature géométrique de la question (direction invariante par l'opérateur) nous montre que les valeurs caractéristiques de $\mathfrak{A}$ sont les mêmes que celles de A et que les solutions caractéristiques seront définies par

$$\Xi_i = S^{-1}X_i.$$

Vérifions-le directement; la relation

$$(\lambda_i - \mathfrak{A})\Xi_i = 0$$

définit la valeur caractéristique λ_i et la solution caractéristique $\Xi_i \neq 0$. Remarquons que λ_i étant une matrice scalaire,

$$\lambda_i = S^{-1}\lambda_i S;$$

nous avons donc

$$(S^{-1}\lambda_i S - S^{-1}AS)\Xi_i = 0.$$

Soit

$$S^{-1}(\lambda_i - A)S\,\Xi_i = 0$$

et, S^{-1} étant une matrice non dégénérée,

$$(\lambda_i - A)S\,\Xi_i = 0.$$

En posant $X_i = S\Xi_i$, il vient

$$(\lambda_i - A)X_i = 0;$$

λ_i est donc une valeur caractéristique de A et $X_i = S\Xi_i$ une solution caractéristique. La réciproque se démontre de la même manière et nous arrivons au théorème annoncé : *la matrice* A *et la matrice* $\mathfrak{A} = S^{-1}AS$ *ont les mêmes valeurs caractéristiques* λ_i; les vecteurs caractéristiques relatifs à ces deux matrices et correspondant à un même λ_i se correspon-

dent par la relation $X = S\Xi$, qui exprime simplement le changement de coordonnées. Cette dernière propriété nous met en évidence l'invariance des ν_i par un changement de coordonnées.

Nous laisserons au lecteur le soin de reprendre cette question, en envisageant, non plus un changement de coordonnées, mais une transformation linéaire régulière de l'espace (autre interprétation fondamentale des matrices S).

4° Envisageons maintenant les puissances positives entières de la matrice A; nous avons par hypothèse :

$$AX_i = \lambda_i X_i.$$

Multiplions à gauche par la matrice A :

$$A^2 X_i = A\lambda_i X_i = \lambda_i AX_i,$$

soit

$$A^2 X_i = \lambda_i^2 X_i.$$

Et si nous recommençons encore cette opération un nombre quelconque de fois, nous arrivons à

$$A^n X_i = \lambda_i^n X_i,$$

ce qui nous démontre que :

Si λ_i est une valeur caractéristique de la matrice A, et X_i une solution propre correspondante, λ_i^n est une valeur caractéristique de la matrice A^n, et X_i est une solution propre correspondante. Mais la réciproque de cette proposition n'est pas toujours vraie.

Cette proposition permet deux extensions :

Remarquons que nous l'avons démontrée pour n entier positif et pour $n = -1$; pour $n = 0$ elle est encore vraie, et c'est là un cas banal.

Nous en déduisons que :

λ_i étant une valeur caractéristique de A non dégénérée et X_i une solution propre correspondante, λ_i^{-1} sera une valeur caractéristique de A^{-1} et X_i une solution propre correspondante; $(\lambda_i^{-1})^n = \lambda_i^{-n}$ sera une valeur caractéristique de $(A^{-1})^n = A^{-n}$, et X_i une solution propre correspondante. La proposition s'étend donc aux puissances entières négatives de la matrice A.

Considérons l'autre extension : soit $P(A) = p_0 + p_1 A + \ldots + p_\nu A^\nu$, un polynome de la matrice A; si nous suppposons

$$\lambda_i X_i = AX_i,$$

nous avons démontré que

$$\lambda_i^k X_i = A^k X_i;$$

il en résulte immédiatement que

$$P(\lambda_i) X_i = P(A) X_i;$$

donc, *si* λ_i *est une valeur caractéristique de la matrice* A *et* X_i *une solution propre correspondante,* $P(\lambda_i)$ *est une valeur caractéristique de la matrice* $P(A)$ *et* X_i *est une solution propre correspondante.* Ici non plus nous n'examinerons pas comment les valeurs caractéristiques de A et ses solutions propres peuvent se déduire de celles de $P(A)$.

2. Résolvante de A. — Si λ n'est pas une valeur caractéristique, le système d'équations non homogènes représenté symboliquement par

$$(\lambda - A) X = Y$$

admet une solution unique X, quel que soit Y, et cette solution est

$$X = (\lambda - A)^{-1} Y.$$

La matrice $(\lambda - A)^{-1}$, fonction de λ, s'appelle la *résolvante* de A et se désigne par la notation

$$(\lambda - A)^{-1} = \mathcal{R}_\lambda = \| R_{ik}(\lambda) \|.$$

D'autres auteurs (Hilbert, Courant) adoptent des notations analogues à celles des équations intégrales; A étant l'analogue du noyau, les μ critiques sont les inverses des λ_i caractéristiques, et le système d'équations

$$(1 - \mu A) X = Y.$$

ou

$$Y = X - \mu A X$$

présente, si μ n'est pas une valeur critique, une solution unique fournie par la relation

$$X = (1 - \mu A)^{-1} Y$$

ou

$$X = Y + \mu R_\mu Y$$

en posant $(1 - \mu A)^{-1} = 1 + \mu R_\mu$ pour exprimer X en Y sous une forme symétrique de celle de Y en X.

C'est R_μ que ces auteurs désignent sous le nom de résolvante. Si A est le noyau de la transformation $1 - \mu A$, R_μ s'appelle le *noyau résolvant.*

Ces deux résolvantes se ramènent d'ailleurs l'une à l'autre par une

relation très simple; posons en effet $\lambda = \frac{1}{\mu}$; l'expression

$$(\lambda - A)X = Y$$

devient alors

$$\left(\frac{1}{\mu} - A\right)X = Y$$

ou bien

$$(1 - \mu A)X = \mu X,$$

soit encore

$$X = \mu(1 - \mu A)^{-1} Y.$$

D'où la relation

$$\mathcal{R}_\lambda Y = \mu(1 - \mu A)^{-1} Y = (\mu + \mu^2 R_\mu)Y, \qquad \text{pour} \quad \lambda = \frac{1}{\mu}$$

quel que soit le vecteur Y; nous avons donc

$$\mathcal{R}_\mu = \mu + \mu^2 R_\mu \qquad \text{avec} \quad \lambda = \frac{1}{\mu}.$$

Démontrons une relation fonctionnelle importante due à Hilbert. A cet effet envisageons deux valeurs λ' et λ'' quelconques, mais non caractéristiques; les deux matrices $\lambda' - A$ et $\lambda'' - A$ sont permutables comme on peut s'en assurer immédiatement ([1]). Si deux matrices sont permutables

$$CD = DC,$$

si de plus elles ne sont pas dégénérées, leurs produits ne le sont pas non plus, et nous pouvons prendre l'inverse des deux membres :

$$D^{-1} C^{-1} = C^{-1} D^{-1};$$

ce qui prouve que leurs inverses sont permutables. Par conséquent, les matrices $\mathcal{R}_{\lambda'}$ et $\mathcal{R}_{\lambda''}$ sont permutables; or nous avons

$$(\lambda' - A)\mathcal{R}_{\lambda'} = 1,$$

c'est-à-dire

$$\lambda' \mathcal{R}_{\lambda'} - 1 = A \mathcal{R}_{\lambda'},$$

et de même

$$\lambda'' \mathcal{R}_{\lambda''} - 1 = A \mathcal{R}_{\lambda''}.$$

Multiplions respectivement ces deux relations par $\mathcal{R}_{\lambda''}$ et $\mathcal{R}_{\lambda'}$ et retranchons; il vient

$$\lambda' \mathcal{R}_{\lambda'} \mathcal{R}_{\lambda''} - \mathcal{R}_{\lambda''} = \lambda'' \mathcal{R}_{\lambda''} \mathcal{R}_{\lambda'} - \mathcal{R}_{\lambda'},$$

([1]) Cette propriété est valable pour tous les polynomes en A, car $P(A)Q(A) = R(A)$, si R est produit des deux polynomes P et Q.

soit

$$(\lambda' - \lambda'')\mathcal{R}_{\lambda'}\mathcal{R}_{\lambda''} = \mathcal{R}_{\lambda''} - \mathcal{R}_{\lambda'}$$

qui est la relation de Hilbert.

3. Étude de la résolvante. — Étudions la résolvante,

$$\mathcal{R}_\lambda = \| R_{ik}(\lambda) \| = (\lambda - A)^{-1}.$$

D'après sa définition, nous avons

$$R_{ik}(\lambda) = \frac{S_{ik}(\lambda)}{f(\lambda)},$$

en désignant par $S_{ik}(\lambda)$ le complément algébrique de $\lambda\delta_{ki} - a_{ki}$ dans le déterminant $f(\lambda)$; S_{ik} est donc un polynome en λ de degré $n-1$. En posant

$$f(\lambda) = (\lambda - \mu_1)^{n_1}(\lambda - \mu_2)^{n_2}\ldots(\lambda - \mu_p)^{n_p},$$

où $\mu_1, \mu_2, \ldots, \mu_p$, sont des racines distinctes, nous obtenons pour R_{ik} une décomposition en fractions simples de la forme

$$R_{ik}(\lambda) = \sum_{j=1}^{p}\sum_{h=1}^{n_j} \frac{r_{ik}^{(hj)}}{(\lambda - \mu_j)^h}.$$

R_{ik} est une fonction holomorphe de λ au voisinage d'une valeur λ différente d'une valeur caractéristique. Les μ_j sont donc les seuls points singuliers *possibles* des R_{ik}; mais *certains* R_{ik} peuvent être réguliers en certains μ_j comme le prouve l'exemple suivant :

Soit la matrice diagonale

$$A = \| \mu_i \delta_{ik} \|;$$

nous avons

$$\lambda - A = \| (\lambda - \mu_i)\delta_{ik} \|$$

et

$$(\lambda - A)^{-1} = \left\| \frac{\delta_{ik}}{\lambda - \mu_i} \right\|;$$

ou encore

$$R_{ik} = \frac{\delta_{ik}}{\lambda - \mu_i};$$

et par conséquent pour $\lambda = \mu_i$, seul le terme R_{ii} est singulier.

Mais nous pouvons assurer qu'en un point du spectre $\lambda = \mu$, il est impossible que tous les R_{ik} soient réguliers; supposons en effet qu'ils le soient tous et choisissons alors une suite $\lambda_1, \lambda_2, \ldots, \lambda_s, \ldots$ de points n'appartenant pas au spectre et tendant vers la valeurs μ du spectre lorsque s augmente indéfiniment. Dans ces conditions, les fonc-

tions $R_{ik}(\lambda_s)$ existent et nous avons

$$(\lambda_s - A)^{-1} \mathcal{R}_{\lambda_s} = 1,$$

soit, sous forme développée,

$$\sum_{l=1}^{n} R_{il}(\lambda_s)[\lambda_s \delta_{lk} - a_{lk}] = \delta_{ik} \quad \text{pour} \quad i, k = 1, 2, \ldots, n.$$

La fonction $R_{il}(\lambda)$ étant supposée régulière en μ, $R_{il}(\lambda_s)$ tend vers $R_{il}(\mu)$ quand s tend vers l'infini et les équations précédentes donnent

$$\sum_{l=1}^{n} R_{il}(\mu)[\mu \delta_{lk} - a_{lk}] = \delta_{ik}.$$

La matrice $(\lambda - A)$ présenterait donc pour $\lambda = \mu$, valeur du spectre, une réciproque $\| R_{ik}(\mu) \|$, ce qui est impossible.

Le spectre est donc *l'ensemble des points singuliers de la matrice* $\mathcal{R}_\lambda = (\lambda - A)^{-1}$, ces singularités étant constituées par des *pôles*; ceci n'est plus vrai pour les matrices infinies, puisqu'il peut, par exemple, exister des spectres continus.

4. Développement de la résolvante. — Nous avons obtenu ci-dessus l'expression

$$R_{ik}(\lambda) = \sum_{j=1}^{n} \sum_{h=1}^{n_j} \frac{r_{ik}^{(hj)}}{(\lambda - \mu_j)^h},$$

somme d'un nombre fini de fractions simples. D'après Laurent, chacune de ces fractions est développable en série de puissances de $\frac{1}{\lambda}$, absolument convergente lorsque le module de λ est supérieur à ρ, *maximum des modules des* μ_j. En dehors du cercle de rayon $\rho + \varepsilon$ (ε positif et arbitrairement petit), nous avons donc le développement suivant absolument et uniformément convergent :

$$R_{ik}(\lambda) = \sum_{m=0}^{\infty} \frac{a_{ik}^{(m)}}{\lambda^{m+1}}.$$

Dans le cercle $|\lambda| < \rho$, une *au moins* de ces séries doit être divergente, car un R_{ik} *au moins* présente un pôle de module ρ. Dans le cas particulier où tous les μ_j sont nuls, ρ est nul, mais le point $\lambda = 0$ est encore un *pôle* et par conséquent tous les $a_{ik}^{(m)}$ doivent être nuls lorsque m surpasse un certain nombre m_0.

Si nous posons

$$\mathrm{A}^{(m)} = \| a_{ik}^{(m)} \|,$$

le développement peut s'écrire en abrégé :

$$\mathcal{R}_\lambda = \sum_{m=0}^{\infty} \frac{\mathrm{A}^{(m)}}{\lambda^{m+1}}$$

(convergent pour $|\lambda| > \rho$), ce qui introduit une série de matrices. Nous pouvons en définir la somme de deux façons qui sont manifestement équivalentes.

a. C'est une matrice dont un élément est la somme de la série formée par les éléments correspondants des matrices de la série.

b. On peut effectuer la somme des k premières matrices, faire tendre k vers l'infini, puis prendre la limite de la matrice somme. Par ce procédé nous devons définir la limite d'une matrice comme étant la matrice dont les éléments sont les limites des éléments correspondants des matrices variables.

Précisons la nature des matrices $\mathrm{A}^{(m)}$. A cet effet, revenons à la définition de la résolvante :

$$(\lambda - \mathrm{A})\,\mathcal{R}_\lambda = 1.$$

Le premier membre est une série entière en $\frac{1}{\lambda}$, absolument et uniformément convergente lorsque $|\lambda| > \rho + \varepsilon$, tandis que le second membre est une série en $\frac{1}{\lambda}$ réduite à 1; identifions, il vient successivement :

$$\mathrm{A}^{(0)} = 1 \qquad (\text{terme constant}),$$

$$\mathrm{A}^{(1)} - \mathrm{A}\mathrm{A}^{(0)} = 0 \qquad \left(\text{terme en } \frac{1}{\lambda}\right),$$

. . . . ,

$$\mathrm{A}^{(m)} - \mathrm{A}\mathrm{A}^{(m-1)} = 0 \qquad \left(\text{terme en } \frac{1}{\lambda^m}\right),$$

. . . .

D'où nous déduisons :

$$\mathrm{A}^{(0)} = 1,$$
$$\mathrm{A}^{(1)} = \mathrm{A},$$
$$\mathrm{A}^{(2)} = \mathrm{A}^2,$$
. . . . ,
$$\mathrm{A}^{(m)} = \mathrm{A}^m,$$
. . . .

Le développement obtenu peut donc s'écrire :

$$\mathcal{R}_\lambda = (\lambda - A)^{-1} = \sum_{m=0}^{\infty} \frac{A^m}{\lambda^{m+1}}.$$

C'est le développement de Liouville-Neumann, convergent lorsque $|\lambda| > \rho + \varepsilon$, ρ désignant le plus grand module du spectre (c'est-à-dire convergent hors du plus petit cercle de centre O contenant tout le spectre).

Il est clair que nous pouvons déduire de cette relation un développement de la résolvante R_μ de Hilbert :

$$\lambda = \frac{1}{\mu}, \qquad \mathcal{R}_\lambda = \mu + \mu^2 R_\mu;$$

or

$$\mathcal{R}_\lambda = \sum_{m=0}^{\infty} A^m \mu^{m+1},$$

d'où

$$R_\mu = A + \mu A^2 + \mu^2 A^3 + \ldots + \mu^m A^{m+1} + \ldots,$$

série convergente dans le plus grand cercle laissant le spectre à l'extérieur.

Le développement de $\mathcal{R}_\lambda$ pouvait être prévu par des opérations formelles :

On a d'abord

$$\mathcal{R}_\lambda = \frac{1}{\lambda - A} = \frac{1}{\lambda}\left[1 + \frac{A}{\lambda} + \ldots + \frac{A^m}{\lambda^m} + \ldots\right] = \frac{1}{\lambda} + \frac{A}{\lambda^2} + \ldots + \frac{A^m}{\lambda^{m+1}} + \ldots.$$

La matrice, ainsi définie, satisfait formellement à la relation

$$(\lambda - A)\mathcal{R}_\lambda = 1.$$

La vérification en est immédiate, ou nous pouvons l'affirmer sans aucun calcul en nous appuyant sur le fait qu'un calcul effectué sur des polynomes d'une variable est encore valable lorsqu'on y remplace cette variable par une matrice. Si nous procédons de cette façon, il nous suffit donc de prouver la convergence de la série de matrices. Pour faire cette démonstration, nous allons utiliser la notion de *valeur absolue de matrices* de Frobenius (1).

5. Avec Frobenius nous désignerons par *valeur absolue de la*

---

(1) *Sitzber.* Berlin, 1911.

matrice A et nous écrirons $|A|$ le nombre positif défini par

$$|A|^2 = \sum_{i,k} |a_{ik}|^2.$$

La valeur absolue est donc la trace (Spur) des deux matrices AA^* et A^*A. Cette notion est très importante pour les matrices finies, mais elle est *défaillante pour les matrices infinies* (excepté celles dites complètement continues). Elle tire son intérêt de son *invariance par les transformations unitaires :*

$$\mathfrak{A} = U^{-1} A U \qquad \text{avec} \quad U^{-1} = U^*;$$

donc

$$\mathfrak{A}^* = U^* A^* U,$$

et, par conséquent,

$$\mathfrak{A}\mathfrak{A}^* = U^{-1} A U U^* A^* U = U^{-1} A A^* U;$$

la propriété envisagée devient alors évidente puisque la trace est invariante dans une transformation linéaire homogène.

Ceci étant posé nous appellerons distance de deux matrices A et B la quantité

$$\Delta(A, B) = |B - A|,$$

et nous aurons évidemment

$$\Delta(A, B) = \Delta(B, A) \geqq 0,$$

l'égalité à zéro ne pouvant avoir lieu que si $A = B$.

Les valeurs absolues de Frobenius vérifient encore les deux propriétés fondamentales suivantes :

a. L'inégalité triangulaire, à savoir

$$|A + B| \leqq |A| + |B|.$$

En effet,

$$|A + B|^2 = \sum_{i,k} |a_{ik} + b_{ik}|^2 \leqq \sum_{i,k} (|a_{ik}| + |b_{ik}|)^2,$$

soit

$$|A + B|^2 \leqq |A|^2 + |B|^2 + 2\sum_{i,k} |a_{ik}|\,|b_{ik}|;$$

et, en utilisant l'inégalité de Cauchy-Schwarz,

$$(\Sigma\, |a_{ik}|\,|b_{ik}|)^2 \leqq \Sigma\, |a_{ik}|^2\, \Sigma\, |b_{ik}|^2.$$

nous obtenons

$$|A + B|^2 \leqq |A|^2 + |B|^2 + 2\,|A|\,|B| = (|A| + |B|)^2,$$

ce qui démontre l'inégalité triangulaire. Celle-ci se généralise à une somme d'un nombre quelconque de termes, et par passage à la limite aux intégrales

$$\left|\int_0^0 A(t)\,dt\right| \leqq \int_0^0 |A(t)|\,dt,$$

en définissant l'intégrale d'une matrice (de même qu'une série de matrices) soit par l'intégrale des termes, soit par une somme finie et un passage à la limite.

b. L'inégalité du produit

$$|AB| \leqq |A|\,|B|,$$

qui, en explicitant, s'écrit

$$\sum_{i,k}\left|\sum_l a_{il}b_{lk}\right|^2 \leqq \sum_{ik}|a_{ik}|^2\sum_{ik}|b_{ik}|^2.$$

Elle se déduit de l'inégalité de Cauchy-Schwarz

$$\left(\sum_l |a_{il}b_{lk}|\right)^2 \leqq \sum_l |a_{il}|^2\sum_l |b_{lk}|^2,$$

que nous sommons en i et k

$$\sum_{i,k}\left(\sum_l |a_{il}b_{lk}|\right)^2 \leqq \left(\sum_{il}|a_{il}|^2\right)\left(\sum_{lk}|b_{lk}|^2\right).$$

L'inégalité est alors démontrée en remarquant que

$$\sum_{i,k}\left|\sum_l a_{il}b_{lk}\right|^2 \leqq \sum_{i,k}\left(\sum_l |a_{il}b_{lk}|\right)^2,$$

puisque

$$\left|\sum_l a_{il}b_{lk}\right| \leqq \sum_l |a_{il}b_{lk}|.$$

6. Revenons maintenant à l'étude de la convergence de la série de Liouville-Neumann

$$\sum_{m=0}^{\infty}\frac{A^m}{\lambda^{m+1}}.$$

D'après la dernière propriété obtenue, nous avons

$$|A^m| \leqq |A|^m.$$

C'est un résultat classique sur les séries géométriques, que la série $\sum \frac{|A|^m}{|\lambda|^{m+1}}$ est convergente lorsque $|\lambda| > |A|$. Si nous remarquons d'autre part que

$$|a_{ik}^{(m)}| \leqq |A^m| \leqq |A|^m,$$

il en résulte que, dans ces conditions, les séries

$$\sum_m \frac{a_{ik}^{(m)}}{\lambda^{m+1}}$$

convergent quels que soient i et k et par conséquent que la série matricielle de Liouville-Neumann converge.

Mais remarquons que ce procédé nous assure seulement de sa convergence pour

$$|\lambda| > |A|,$$

et ne nous indique pas d'une façon précise *toute la région* de convergence. En comparant ce résultat et celui obtenu dans le numéro 4, nous trouvons que, nécessairement, ρ est inférieur ou égal à $|A|$, résultat qu'il est très facile de vérifier directement en revenant à la définition de $|A|$:

$$|A|^2 = \text{trace normale de } A = \text{trace de } AA^* = \text{trace de } A^*A.$$

D'autre part, nous avons montré que $|A|$ est invariant par une *transformation unitaire*. Nous allons profiter de cette latitude pour donner à A une forme particulièrement simple. Nous verrons plus loin que nous pouvons, par une *transformation unitaire*, ramener A à la forme de Schur :

$$A = \begin{Vmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ .. & \lambda_2 & \dots & 0 & 0 \\ & & \ddots & & \\ .. & .. & \dots & \lambda_{n-1} & 0 \\ .. & .. & \dots & \dots & \lambda_n \end{Vmatrix}$$

les termes situés au-dessus de la diagonale principale étant tous nuls, et les termes de la diagonale principale étant égaux aux valeurs caractéristiques; il est clair alors que

$$|A|^2 = \text{trace de } AA^* = \sum_{i=1}^{n} \lambda_i^2.$$

Cette relation entraîne

$$|A|^2 \geqq \rho^2,$$

si ρ désigne les λ_i maximum; c'est ce que nous voulions établir.

7. Donnons, de la convergence de la série de Neumann, une autre démonstration qui aura sur la précédente l'avantage de s'étendre aux matrices infinies dans des cas plus généraux. Envisageons dans la matrice A^m la somme des valeurs absolues des éléments de chaque ligne

$$\theta_i^{(m)} = \sum_k |a_{ik}^{(m)}|,$$

et désignons par M un nombre supérieur à tous les $\theta_i^{(1)}$ quel que soit i; nous allons prouver par récurrence que

$$\underline{\theta_i^{(m)} < M^m}, \qquad \text{d'où résultera } |a_{ik}^{(m)}| < M^m.$$

En effet,

$$a_{ik}^{(m)} = \sum_j a_{ij}^{(m-1)} a_{jk}^{(1)},$$

d'où

$$|a_{ik}^{(m)}| \leqq \sum_j |a_{ij}^{(m-1)}|\,|a_{jk}^{(1)}|$$

et

$$\theta_i^{(m)} \leqq \sum_{j,k} |a_{ij}^{(m-1)}|\,|a_{jk}^{(1)}| = \sum_j \left[|a_{ij}^{(m-1)}| \sum_k |a_{jk}^{(1)}| \right].$$

Si nous supposons que la proposition est vraie jusque $m-1$,

$$\theta_i^{(m)} < \sum_j |a_{ij}^{(m-1)}|\,M < M M^{m-1} = M^m,$$

nous démontrons qu'elle est encore vraie pour m. Or elle est vraie pour $m = 1$, elle l'est donc pour toutes les valeurs de m.

En remarquant que

$$\left| \frac{a_{ik}^{(m)}}{\lambda^{m+1}} \right| < \frac{M^m}{|\lambda|^{m+1}},$$

nous voyons immédiatement que la série de Neumann est convergente pour $|\lambda| > M$, la série relative à chaque élément se trouvant, dans ces conditions, majorée par une série géométrique convergente.

Cette méthode a subi diverses modifications; en particulier, pour le cas où

$$\sum_k |a_{ik}|^2 < M^2,$$

quel que soit i, ce qui permet l'application aux matrices infinies bornées.

8. Application de ces évaluations. — *Séries entières de matrices.* — Nous avons montré que $R_{ik}(\lambda)$ était défini par la série convergente $\sum_m \frac{a_{ik}^{(m)}}{\lambda^{m+1}}$ lorsque $|\lambda| > \rho$, *ρ étant la borne supérieure du spectre.* Donc l'expression $\frac{|a_{ik}^{(m)}|}{(\rho+\varepsilon)^{m+1}}$ tend vers zéro lorsque m augmente indéfiniment si ε est un nombre positif quelconque. Par conséquent, il est possible d'associer à ε un nombre fixe $\alpha(\varepsilon)$ tel que

$$|a_{ik}^{(m)}| < \alpha(\varepsilon)(\rho+\varepsilon)^{m+1}$$

quel que soit m. D'autre part, considérons une série entière

$$g(z) = \sum_0^\infty g_p z^p,$$

ayant R pour rayon de convergence, et formons

$$g_m(A) = \sum_0^m g_p A^p.$$

C'est une matrice dont l'élément général est $\sum_0^m g_p a_{ik}^{(p)}$. Pour que $g_m(A)$ ait une limite lorsque m augmente indéfiniment, il faut et il suffit (par définition même) que les séries

$$\sum_0^\infty g_p a_{ik}^{(p)}$$

soient convergentes quels que soient i et k. Or, d'après ce que nous venons de voir, ces séries sont majorées par

$$\alpha(\varepsilon)\sum_0^\infty |g_p|(\rho+\varepsilon)^{p+1} = \alpha(\varepsilon)(\rho+\varepsilon)\sum_0^\infty |g_p|(\rho+\varepsilon)^p,$$

convergente si $\rho+\varepsilon < R$, c'est-à-dire si $\rho < R$, ou bien encore : *si le cercle de convergence de la série entière contient le spectre à son intérieur.*

En résumé, *la série* $\sum_0^\infty g_m A^m$ *converge absolument et définit la matrice* $g(A)$ *si le cercle de convergence de la série entière* $\sum_0^\infty g_m z^m$ *contient à son intérieur le spectre de la matrice* A.

En particulier, cela nous permet de définir

$$e^{A} = \sum_{0}^{\infty} \frac{A^m}{m\,!},$$

$$\cos A = \sum_{0}^{\infty} \frac{(-1)^m A^{2m}}{(2m)!},$$

$$\sin A = \sum_{0}^{\infty} \frac{(-1)^m A^{2m+1}}{(2m+1)!},$$

ces séries étant toujours absolument convergentes. Les théorèmes classiques d'addition et de multiplication des fonctions exponentielles et circulaires sont encore vrais ici si l'on considère des *matrices permutables*; car les calculs les vérifiant à l'aide des séries sont formellement les mêmes, que l'on opère sur des variables ordinaires ou sur des matrices permutables. Par exemple, si A et B sont permutables, nous avons

$$e^{A} e^{B} = e^{A+B}.$$

Cas particulier :

$$e^{At} e^{At'} = e^{A(t+t')}.$$

Autre application : Résolution du système différentiel

$$\frac{dx}{dt} = \sum_{k=1}^{n} a_{ik} x_k \qquad \text{pour} \quad i = 1, 2, \ldots, n.$$

Nous pouvons poser le problème sous forme vectorielle

$$\frac{dX}{dt} = AX,$$

X étant le vecteur $(x_1, x_2, \ldots, x_n)$ avec les conditions initiales $X(0) = X_0$. Et nous allons montrer que la solution est

$$X(t) = e^{At} X_0.$$

En effet, $X(t)$ s'obtient à partir de X_0 par une matrice dépendant du temps :

$$X(t) = B(t) X_0,$$

puisque l'équation différentielle proposée est linéaire en X. Supposons que la matrice $B(t)$ se développe en série entière de la variable t :

$$B(t) = \sum_{m=0}^{\infty} \frac{A_m t^m}{m\,!},$$

les A_m étant des matrices d'ordre n. En reportant dans l'équation différentielle et identifiant, il vient

$$\begin{aligned} A_0 &= 1, \\ A_1 &= AA_0, \\ A_2 &= AA_1, \\ &\ldots\ldots\ldots\ldots, \\ A_m &= AA_{m-1}, \\ &\ldots\ldots\ldots\ldots\ldots, \end{aligned}$$

soit $A_m = A^m$; la solution cherchée est donc

$$A(t) = \sum_{m=0}^{\infty} \frac{A^m t^m}{m!} = e^{At},$$

série toujours convergente et l'on a bien

$$X(t) = X_0 e^{At} \quad (^1).$$

V. — FORME BILINÉAIRE ASSOCIÉE A UNE MATRICE.

1. Étant donnée une matrice $A = \| a_{ik} \|$, nous lui ferons correspondre la forme bilinéaire

$$\sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} x_i y_k,$$

qu'avec Hilbert nous désignerons par $A(X, Y)$, X étant le vecteur de coordonnées x_i et Y celui de coordonnées y_i. Signalons en passant que d'autres auteurs, Wintner en particulier, utilisent, pour désigner la forme bilinéaire, la notation $\psi(A; X, Y)$.

Remarquons qu'à une matrice correspond une forme bilinéaire et que, réciproquement, à une forme bilinéaire correspond une matrice et une seule.

En langage matriciel, $\sum_k a_{ik} y_k$ est un vecteur qui se note AY, et la forme bilinéaire

$$A(X, Y) = X'AY,$$

X' étant une matrice à une ligne, transposée de la matrice X à une colonne.

---

(1) Pour d'autres applications, *voir* WEYL, *Quantentheorie*, Chap. I, § 6.

En particulier, à la matrice unité correspond la forme bilinéaire unité

$$E(X, Y) = X'Y = \sum_{i=1}^{n} x_i y_i.$$

Considérons la relation

$$(1) \qquad Y = AX;$$

si nous désignons par U un *vecteur indéterminé*, cette relation est équivalente à

$$U'Y = U'AX,$$

soit, avec la notation des formes bilinéaires,

$$(2) \qquad E(U, Y) = A(U, X).$$

Il est équivalent de dire que les deux vecteurs X et Y vérifient la relation (1) ou satisfont identiquement en U à la relation (2). Ce résultat peut aussi s'obtenir en parlant formes et équations sans utiliser les vecteurs :

(1) se trouve alors remplacé par le système

$$y_i = \sum_{k=1}^{n} a_{ik} x_k \quad \text{pour} \quad i = 1, 2, \ldots, n,$$

qui est équivalent à *l'identité en u* suivante :

$$\sum_{i=1}^{n} u_i y_i = \sum_{i,k=1}^{n} a_{ik} u_i x_k,$$

qui n'est autre que le développement de la relation (2).

Nous appellerons *produit symbolique* (allemand : Faltung) C(U, X) des deux formes bilinéaires A(U, X) et B(U, X) et nous noterons

$$C(U, X) = A(U, X)\, B(U, X),$$

ou plutôt

$$C(U, X) = A(U, \bullet)\, B(\bullet, X),$$

la forme bilinéaire

$$\sum_{i,k=1}^{n} c_{ik} u_i x_k,$$

avec

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

En d'autres termes, le produit symbolique est la forme bilinéaire associée à la matrice produit des matrices A et B; nous pouvons donner à ce produit symbolique une forme intéressante :

$$\mathrm{C}(\mathrm{U}, \mathrm{X}) = \sum_{ik}\sum_{j} a_{ij} b_{jk} u_i x_k = \sum_{j}\left[\sum_{k} b_{jk} x_k\right]\left[\sum_{i} a_{ij} u_i\right];$$

or, nous avons

$$\sum_{i=1}^{n} a_{ij} u_i = \frac{\partial \mathrm{A}(\mathrm{U}, \mathrm{X})}{\partial x_j}.$$

et de même

$$\sum_{k=1}^{n} b_{jk} x_k = \frac{\partial \mathrm{B}(\mathrm{U}, \mathrm{X})}{\partial u_j}.$$

D'où l'expression cherchée

$$\mathrm{C}(\mathrm{U}, \mathrm{X}) = \sum_{j=1}^{n} \frac{\partial \mathrm{A}}{\partial x_j} \frac{\partial \mathrm{B}}{\partial u_j}.$$

Nous pouvons encore vérifier ce résultat de la façon suivante :

$\frac{\partial \mathrm{A}}{\partial x_j}$ est la $j^{\text{ième}}$ composante du vecteur $\mathrm{A'U}$,

$\frac{\partial \mathrm{B}}{\partial u_j}$ est la $j^{\text{ième}}$ composante du vecteur BX,

et par conséquent

$$\sum_{j=1}^{n} \frac{\partial \mathrm{A}}{\partial x_j} \frac{\partial \mathrm{B}}{\partial u_j} = (\mathrm{A'U})'\mathrm{BX} = \mathrm{U'ABX},$$

qui est bien la forme bilinéaire associée à la matrice $\mathrm{C} = \mathrm{AB}$.

Ayant défini le produit de deux formes bilinéaires, nous avons défini le carré d'une forme; nous en déduisons le cube, et ainsi de suite jusqu'à la $n^{\text{ième}}$ puissance que nous appellerons $n^{\text{ième}}$ *itérée de* $\mathrm{A}(\mathrm{U}, \mathrm{X})$, et que nous noterons $\mathrm{A}^n(\mathrm{U}, \mathrm{X})$. Ce n'est autre que la forme bilinéaire associée à A^n.

De même, nous définirons la *forme réciproque* d'une forme bilinéaire associée à une matrice A non dégénérée, la *forme associée à la matrice réciproque* A^{-1}. En désignant par A_{ik} le complément algébrique de a_{ik} dans le déterminant Δ de la matrice A, nous avons montré que

$$\mathrm{A}^{-1} = \left\| \frac{\mathrm{A}_{ki}}{\Delta} \right\|.$$

La forme réciproque s'écrit donc :

$$\mathrm{A}^{-1}(\mathrm{U}, \mathrm{X}) = \sum_{i,k} \frac{\mathrm{A}_{ki}}{\Delta} u_i x_k = \begin{vmatrix} 0 & u_1 & u_2 & \dots & u_n \\ x_1 & a_{11} & a_{12} & \dots & a_{1n} \\ x_2 & a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_n & a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \times \left(-\frac{1}{\Delta}\right).$$

Étudions maintenant l'*effet d'un changement de variables* sur les vecteurs X et Y ; nous avons

$$\mathrm{A}(\mathrm{X}, \mathrm{Y}) = \mathrm{X}'\mathrm{A}\mathrm{Y},$$

et les changements de variables s'expriment par les deux matrices régulières B et C :

$$\mathrm{X} = \mathrm{B}\Xi,$$
$$\mathrm{Y} = \mathrm{C}\mathrm{H}.$$

Dans ces conditions

$$\mathrm{X}'\mathrm{A}\mathrm{Y} = \Xi'\mathrm{B}'\mathrm{A}\mathrm{C}\mathrm{H} = \Xi'\mathfrak{A}\mathrm{H} = \mathfrak{A}(\Xi, \mathrm{H}).$$

La forme bilinéaire transformée sera donc associée à $\mathfrak{A} = \mathrm{B}'\mathrm{A}\mathrm{C}$, *dont les éléments sont* $\alpha_{ik} = \sum_{j,l} b_{ji} a_{jl} c_{lk}$; d'où nous déduisons le développement de la forme bilinéaire transformée

$$\sum_{ik} \xi_i \eta_k \sum_{j,l} b_{ji} a_{jl} c_{lk},$$

que nous aurions pu obtenir directement, sans utiliser les notations matricielles.

2. Nous appellerons *spectre de la forme* A(U, X) le spectre de la matrice A, c'est-à-dire l'ensemble des racines λ du déterminant de la forme bilinéaire

$$\lambda \mathrm{E}(\mathrm{U}, \mathrm{X}) - \mathrm{A}(\mathrm{U}, \mathrm{X}).$$

La résolvante de cette forme est $\mathcal{R}_\lambda(\mathrm{U}, \mathrm{X})$ qui satisfait à

$$[\lambda \mathrm{E}(\mathrm{U}, \mathrm{X}) - \mathrm{A}(\mathrm{U}, \mathrm{X})] \mid \mathcal{R}_\lambda(\mathrm{U}, \mathrm{X}) = \mathrm{E}(\mathrm{U}, \mathrm{X}).$$

Dans la théorie des équations intégrales, on considère, en général, comme résolvante une expression un peu différente, que l'on obtient en remplaçant λ par $\frac{1}{\mu}\cdot$

$$\mathrm{Y} = (1 - \mathrm{A}\mu)\mathrm{X} = \mathrm{X} - \mathrm{A}\mu\mathrm{X}$$

s'écrit, dans le langage des formes bilinéaires, comme étant l'identité en U suivante :

$$E(U, Y) = E(U, X) - \mu A(U, X).$$

Nous résolvons par une formule symétrique de la première

$$X = Y + \mu R_\mu Y,$$

qui se traduit par l'identité en U

$$E(U, X) = E(U, Y) + \mu R_\mu(U, Y).$$

La forme associée à la matrice R_μ, c'est-à-dire

$$R_\mu(U, Y) = U' R_\mu Y,$$

est dite *forme résolvante* de la forme $A(U, X)$. En comparant les équations donnant X et Y, nous obtenons immédiatement

$$1 + \mu R_\mu = (1 - \mu A)^{-1},$$

c'est-à-dire

$$R_\mu = \frac{(1 - \mu A)^{-1} - 1}{\mu}.$$

Or nous avons démontré que, lorsque μ est inférieur à $\frac{1}{\rho}$, on peut développer $(1 - \mu A)^{-1}$ par la formule

$$(1 - \mu A)^{-1} = 1 + \mu A + \mu^2 A^2 + \ldots.$$

Il en résulte que, dans ces mêmes conditions,

$$R_\mu = A + \mu A^2 + \ldots + \mu^n A^{n+1} + \ldots;$$

les formes correspondant aux puissances de A étant les itérées de la forme bilinéaire associée à A, il est clair que nous avons ([1])

$$R_\mu(U, Y) = A(U, Y) + \mu A^2(U, Y) + \ldots + \mu^n A^{n+1}(U, Y) + \ldots$$

Cette expression est convergente pour $\mu < \rho^{-1}$. On pose quelquefois

$$\mathcal{R}_\lambda(U, X) = \mathcal{R}(U, X; \lambda)$$

et

$$R_\mu(U, Y) = R(U, Y; \mu).$$

Fredholm a donné une expression de R_μ qui est valable quel que

([1]) Il est à peine besoin de remarquer que la forme associée à une somme de matrices est la somme des formes associées à chaque matrice; le passage à la limite (série) se fait sans aucune difficulté.

soit μ. Nous allons *esquisser* la méthode de Fredholm. La forme bilinéaire associée à $(1-\mu A)^{-1}$ est

$$U'(1-\mu A)^{-1}Y = \begin{vmatrix} 0 & u_1 & u_2 & \dots & u_n \\ y_1 & & & & \\ y_2 & & & & \\ \dots & & \Delta(\mu) & & \\ y_n & & & & \end{vmatrix} \left(-\frac{1}{\Delta(\mu)}\right)$$

en désignant par $\Delta(\mu)$ le déterminant de $1-\mu A$. Si, d'autre part, nous posons la forme

$$\Delta(U, Y; \mu) = \begin{vmatrix} 0 & u_1 & u_2 & \dots & u_n \\ y_1 & & & & \\ y_2 & & & & \\ \dots & & \Delta(\mu) & & \\ y_n & & & & \end{vmatrix},$$

nous avons

$$U'(1-\mu A)^{-1}Y = -\frac{\Delta(U, Y; \mu)}{\Delta(\mu)},$$

et par conséquent

$$R_\mu = -\frac{\Delta(U, Y; \mu)}{\mu\Delta(\mu)} - \frac{1}{\mu}E(U, Y).$$

Les deux déterminants ainsi considérés sont des polynomes en μ de degrés n et $n-1$ que nous pouvons développer

$$\Delta(U, Y; \mu) = \Delta_1(U, Y) - \mu\Delta_2(U, Y) + \mu^2\Delta_3(U, Y) - \dots + (-1)^{n-1}\mu^{n-1}\Delta_n(U, Y),$$

$$\Delta(\mu) = 1 - \mu\Delta_1 + \mu^2\Delta_2 - \dots + (-1)^n\mu^n\Delta_n,$$

en posant

$$\Delta_i(U, Y) = \sum \begin{vmatrix} 0 & u_{k_1} & u_{k_2} & \dots & u_{k_i} \\ y_{k_1} & a_{k_1k_1} & a_{k_1k_2} & \dots & a_{k_1k_i} \\ y_{k_2} & a_{k_2k_1} & a_{k_2k_2} & \dots & a_{k_2k_i} \\ \dots & \dots & \dots & \dots & \dots \\ y_{k_i} & a_{k_ik_1} & a_{k_ik_2} & \dots & a_{k_ik_i} \end{vmatrix},$$

la sommation étant effectuée pour tous les entiers $k_1, k_2, \dots, k_i$ de 1 à n et tels que $k_1 < k_2 < \dots < k_i$; et de même

$$\Delta_i = \sum \begin{vmatrix} a_{k_1k_1} & a_{k_1k_2} & \dots & a_{k_1k_i} \\ a_{k_2k_1} & a_{k_2k_2} & \dots & a_{k_2k_i} \\ \dots & \dots & \dots & \dots \\ a_{k_ik_1} & a_{k_ik_2} & \dots & a_{k_ik_i} \end{vmatrix}.$$

Et nous obtenons ainsi l'expression de Fredholm de la résolvante.

3. Après ces quelques indications nous voyons l'utilisation de la forme A(X, Y). Une matrice A est un ensemble de n^2 nombres que l'on peut déterminer par n^2 équations obtenues en donnant à X n valeurs et à Y n valeurs convenables dans la forme bilinéaire associée A(X, Y). Ceci permet de *raisonner* et de *calculer sur la matrice* A par l'intermédiaire des *valeurs prises par la forme bilinéaire associée.*

VI. — ESPACE DUAL (OU DUALISTIQUE) ASSOCIÉ A L'ESPACE VECTORIEL E_n.

1. Nous avons déjà montré qu'une forme linéaire par rapport aux coordonnées d'un vecteur X peut être définie comme un nombre L attaché à chaque vecteur X, satisfaisant aux deux relations de linéarité

$$L(\alpha X) = \alpha L(X),$$

α désignant un nombre quelconque et X un vecteur quelconque, et

$$L(X_1 + X_2) = L(X_1) + L(X_2),$$

X_1 et X_2 désignant deux vecteurs quelconques. Ces deux propriétés caractérisent une forme linéaire

$$L(X) = \sum_{i=1}^{n} \alpha_i x_i.$$

Effectuons sur le vecteur X un changement de coordonnées, défini par

$$X = A\mathfrak{X};$$

la forme L(X) deviendra alors une certaine forme $\mathfrak{L}(\mathfrak{X})$ satisfaisant à

$$\mathfrak{L}(\mathfrak{X}) = L(X),$$

soit

$$\sum_i \alpha'_i x'_i = \sum_i \alpha_i x_i,$$

lorsque

$$x_i = \sum_k a_{ik} x'_k \qquad \text{pour} \quad i = 1, 2, \ldots, n,$$

Nous devons donc avoir

$$\sum_i \alpha'_i x'_i = \sum_{ik} \alpha_i a_{ik} x'_k = \sum_i x'_i \sum_k \alpha_k a_{ki};$$

ce qui nous montre que les α'_i sont définis à partir des α_i par la relation

$$\alpha'_i = \sum_k a_{ki} \alpha_k.$$

Lorsqu'on effectue un changement de coordonnées sur les x_n, il en résulte sur les coefficients α_i une substitution linéaire *contragrédiente* (contravariante) à celle qui s'exerce sur les composantes x_i des vecteurs de l'espace E_n. Nous dirons que les α_i définissent un vecteur Ξ dans l'*espace dualistique* $\mathcal{E}_n$ associé à l'espace vectoriel E_n. Nous pouvons encore dire que les α_i sont les coordonnées tangentielles du *plan*

$$L(X) = 0.$$

Les deux vecteurs Ξ *et* X, X *quelconque dans* E_n *et* Ξ *quelconque dans l'espace dual associé* $\mathcal{E}_n$, *sont tels que leur produit intérieur est invariant par un changement de coordonnées*, puisque

$$\Xi' X = \xi_1 x_1 + \xi_2 x_2 + \ldots + \xi_n x_n,$$

en désignant dorénavant par ξ_i les composantes du vecteur Ξ.

Nous avons déjà remarqué qu'un changement de coordonnées dans les espaces E_n et $\mathcal{E}_n$ s'exprime par des *substitutions contragrédientes;* vérifions-le sous forme vectorielle

$$X = A \mathfrak{X},$$
$$\Xi = \mathfrak{A} \xi,$$

les matrices A et $\mathfrak{A}$ devant être régulières et telles que

$$\Xi' X = \xi' \mathfrak{X},$$

soit

$$\xi' \mathfrak{A}' A \mathfrak{X} = \xi' \mathfrak{X},$$

quels que soient ξ et $\mathfrak{X}$. Ce qui entraîne la condition

$$\mathfrak{A}' A = 1,$$

soit

$$\mathfrak{A} = (A^{-1})' = (A')^{-1},$$

qui exprime la contragrédience des deux substitutions.

Dans le cas particulier où les deux vecteurs X et Ξ satisfont à la relation $\Xi' X = 0$, ils sont dits en *involution*. L'ensemble des vecteurs Ξ en involution avec un vecteur X forme un plan dans l'espace dual $\mathcal{E}_n$. Nous pouvons donc affirmer qu'un vecteur de E_n détermine un plan de $\mathcal{E}_n$ et réciproquement.

2. Étant donnée une matrice A à n lignes et m colonnes

$$A = \| a_{ik} \| \quad \text{avec} \quad i = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, m,$$

nous avons déjà défini sa matrice *duale* ou *transposée* à m lignes et n colonnes

$$A' = \| a'_{ik} \| \quad \text{avec} \quad a'_{ik} = a_{ki}.$$

L'introduction de l'espace dual va nous permettre de *donner une interprétation de la matrice transposée.* La matrice A définit une transformation ponctuelle permettant de passer d'un espace E_m à un espace E_n par

$$Y = AX, \qquad X = \begin{vmatrix} x_1 \\ \cdot\cdot \\ x_m \end{vmatrix}, \qquad Y = \begin{vmatrix} y_1 \\ \cdot\cdot \\ y_n \end{vmatrix}$$

soit

$$y_i = \sum_{k=1}^{m} a_{ik} x_k \qquad \text{pour} \quad i = 1, 2, \ldots, n.$$

Soient $\mathcal{E}_m$ et $\mathcal{E}_n$ les espaces duals correspondant aux espaces E_m et E_n. De même que X a été choisi arbitrairement dans E_m, nous choisirons arbitrairement H dans $\mathcal{E}_n$. Au vecteur Y de E_n nous ferons correspondre le vecteur Ξ de $\mathcal{E}_m$ tel que soit invariant le produit intérieur

$$H'Y = \Xi'X,$$

quel que soit X; ce qui entraîne

$$H'A = \Xi',$$

ou encore

$$\Xi = A'H.$$

Nous pouvons donc dire que *toute transformation linéaire de* E_m *en* E_n, *soit* $E_n = AE_m$ (*ou* $Y = AX$), *laissant invariante la forme bilinéaire* H'AX *formée avec un vecteur* H *arbitraire de* $\mathcal{E}_n$ *et un vecteur* X *arbitraire de* E_m, *induit une transformation linéaire faisant passer de* $\mathcal{E}_n$ *à* $\mathcal{E}_m$ *suivant la relation*

$$\Xi = A'H,$$

qui est la transposée de

$$Y = AX.$$

Nous dirons quelquefois que A' est l'opérateur dual de A. Remarquons que

$$H'AX = \Xi'X = X'\Xi = X'A'H.$$

3. Décomposition de $\mathcal{E}_n$ corrélative à une décomposition de E_n. — Si E_n est décomposé en deux sous-espaces complémentaires E_1 et E' ayant respectivement comme nombre de dimensions n_1 et n' (avec $n = n_1 + n'$), nous envisageons dans l'espace $\mathcal{E}_n$ dual de E_n *tous les vecteurs* Ξ *qui sont en involution avec tous les vecteurs* X *du sous-espace* E_1. Il est clair que la condition nécessaire et suffisante pour qu'un vecteur Ξ vérifie cette condition est qu'il soit en involution avec les n_1 vecteurs de base de la multiplicité E_1 ; donc qu'il satisfasse à

$$X_1'^{(i)}\Xi = 0 \qquad \text{pour} \quad i = 1, 2, \ldots, n_1.$$

Le vecteur Ξ devant satisfaire à n_1 équations linéaires, homogènes, indépendantes, engendre dans $\mathcal{E}_n$ une multiplicité à $n - n_1$ dimensions que nous désignerons par $\mathcal{E}_1$. D'après la définition même de cette multiplicité, *tout vecteur de* E_1 *est en involution avec tout vecteur de* $\mathcal{E}_1$.

Si un opérateur A de l'espace E_n laisse *invariant le sous-espace* E_1, il en résulte que l'opérateur A' dual de l'espace $\mathcal{E}_n$ laisse *invariant le sous-espace* $\mathcal{E}_1$. En effet : si X_1 appartient à E_1, le vecteur $Y_1 = AX_1$ appartient également à E_1. Nous allons montrer que, de même, si H appartient à $\mathcal{E}_1$, le vecteur $\Xi = A'H$ appartient également à $\mathcal{E}_1$. Le vecteur H appartient à $\mathcal{E}_1$ et le vecteur Y_1 appartient à E_1 ; nous avons donc

$$H'Y_1 = 0.$$

Or $\Xi' = H'A$; par conséquent

$$\Xi' X_1 = H'AX_1 = H'Y_1 = 0,$$

quel que soit le vecteur X_1 de E_1 ; par conséquent le vecteur Ξ appartient à $\mathcal{E}_1$.

Généralisons ce résultat au cas où E se trouve décomposé en un certain nombre de sous-espaces complémentaires E_1, E_2, E_3, ..., ayant respectivement pour nombres de dimensions n_1, n_2, n_3, ...

$$E = E_1 + E_2 + E_3 + \ldots,$$
$$n = n_1 + n_2 + n_3 + \ldots.$$

Considérons, parmi les vecteurs Ξ de $\mathcal{E}$, ceux Ξ_1 qui sont en involution avec tous les vecteurs du sous-espace $E_2 + E_3 + \ldots$; d'après l'étude précédente, ils engendrent un sous-espace $\mathcal{E}_1$ à n_1 dimensions. De même, considérons parmi les vecteurs Ξ de $\mathcal{E}$, ceux Ξ_2 qui sont en involution avec tous les vecteurs du sous-espace $E_1 + E_3 + \ldots$; ils engendrent un sous-espace $\mathcal{E}_2$ à n_2 dimensions, et ainsi de suite.

Il résulte immédiatement de cette définition que, réciproquement, un vecteur quelconque de E_1 est en involution avec un vecteur quelconque de $\mathcal{E}_2 + \mathcal{E}_3 + \ldots$; un vecteur quelconque de E_2 est en involution avec un vecteur quelconque de $\mathcal{E}_1 + \mathcal{E}_3 + \ldots$, etc.

Nous allons montrer que dans $\mathcal{E}$ les sous-espaces $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, ... sont disjoints et complémentaires :

a. Ils sont *disjoints*; à cet effet prouvons que le vecteur

$$\Xi_1 + \Xi_2 + \Xi_3 + \ldots$$

ne peut être nul que si Ξ_1, Ξ_2, Ξ_3, ..., le sont. En effet,

$$\Xi_1 = -(\Xi_2 + \Xi_3 + \ldots)$$

appartenant à $\mathcal{E}_1$ doit être en involution avec les vecteurs de $E_2 + E_3 + \ldots$ et appartenant à $\mathcal{E}_2 + \mathcal{E}_3 + \ldots$ doit aussi être en involution avec tous les vecteurs de E_1. Ξ_1 devant être en involution avec tous les vecteurs de E est nécessairement nul; il en est de même des autres vecteurs Ξ_2, Ξ_3, ... Les sous-espaces $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, ... sont donc disjoints.

b. Ils sont *complémentaires*, car la somme des dimensions de ces espaces *disjoints* est

$$n = n_1 + n_2 + n_3 + \ldots.$$

Nous avons donc bien

$$\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 + \ldots.$$

Nous pouvons dire en un certain sens que $\mathcal{E}_i$ *est l'espace dual de* E_i, car si X_i appartient à E_i et $H = H_1 + H_2 + H_3 + \ldots$ appartient à $\mathcal{E}$ (H_1, H_2, H_3, ... appartenant respectivement à $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ ...), nous avons évidemment

$$X'_i H = X'_i H_i,$$

vu que $X'_i\, H_j$ est nul lorsque i est différent de j, E_i étant dans ces conditions en involution avec $\mathcal{E}_j$.

Tout ceci est fort important pour la métrique, où l'involution deviendra l'orthogonalité.

CHAPITRE III.

DÉFINITION DE L'ESPACE MÉTRIQUE. PREMIÈRES APPLICATIONS. OPÉRATEURS REMARQUABLES DANS L'ESPACE MÉTRIQUE.

I. — LA MÉTRIQUE DANS L'ESPACE VECTORIEL A n DIMENSIONS.

1. Nous allons introduire la métrique dans la géométrie linéaire homogène par la définition de la *longueur* $|X|$ d'un vecteur X. Nous déterminerons ensuite le groupe de transformations linéaires laissant invariante la longueur; nous l'appellerons *groupe unitaire* et nous désignerons par *géométrie unitaire* la géométrie l'ayant pour groupe fondamental.

Dans la géométrie euclidienne, la longueur d'un vecteur X, de coordonnées $x_1, x_2, \ldots, n_n$, est déterminée par la racine carrée de la *forme quadratique fondamentale dite forme unité :*

$$|X| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}.$$

Cette expression est invariante lorsqu'on fait un changement de coordonnées rectangulaires; le groupe de transformations linéaires homogènes ainsi défini, qui n'est autre que le groupe des déplacements, est le groupe fondamental de la géométrie euclidienne.

Cette définition de la longueur est intéressante dans le corps des vecteurs X réels; car la condition nécessaire et *suffisante* pour qu'un vecteur réel soit nul est que sa longueur soit nulle. Cette propriété n'est plus vraie dans le corps des X complexes, la longueur d'un vecteur pouvant être nulle sans que ce vecteur le soit. Alors, nous définirons la longueur d'un vecteur X de coordonnées $x_1, x_2, \ldots, x_n$, *par la forme d'Hermite unité*

$$|X|^2 = \overline{x}_1 x_1 + \overline{x}_2 x_2 + \ldots, + \overline{x}_n x_n.$$

$|X|^2$ s'appelle aussi *norme de* X.

Nous verrons ci-dessous que cette expression n'est autre que le pro-

duit scalaire du vecteur X par lui-même; remarquons que cette forme d'Hermite a toujours une valeur réelle, et que dans le cas d'un vecteur réel elle se réduit à la forme classique de la géométrie euclidienne. Lorsque, relativement à un système de coordonnées, la longueur aura cette forme, nous dirons que le système de coordonnées est *normal*.

Nous avons déjà défini (Chap. II, § II, n° 10) le *produit intérieur* de deux vecteurs; parallèlement au changement de définition de la longueur d'un vecteur, nous allons modifier cette expression de façon que l'on ait comme cas particulier

$$|X|^2 = (X, X),$$

et la nouvelle expression obtenue s'appellera *produit scalaire*. Le *produit scalaire de deux vecteurs* X *et* Y *ayant respectivement pour coordonnées* $x_1, x_2, \ldots, x_n$ *d'une part et* $y_1, y_2, \ldots, y_n$ *d'autre part, est*

$$(X, Y) = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \ldots + \bar{x}_n y_n,$$

si nous nous plaçons dans un système de coordonnées normal.

En notation matricielle on a

$$(X, Y) = \bar{X}' Y = X^* Y.$$

Le produit scalaire, n'est *pas commutatif*, car si nous permutons X et Y nous obtenons l'expression imaginaire conjuguée

$$(Y, X) = \overline{(X, Y)};$$

mais il est distributif relativement à la somme des vecteurs

$$(X' + X'', Y) = (X', Y) + (X'', Y),$$

et de même

$$(X, Y' + Y'') = (X, Y') + (X, Y'').$$

Si l'on multiplie l'un des deux vecteurs par un scalaire a, le produit scalaire est multiplié par $\bar{a}$ ou a suivant que c'est le premier ou le second vecteur du produit qui est multiplié par a :

$$(aX, Y) = \bar{a}(X, Y),$$
$$(X, aY) = a(X, Y);$$

et nous en déduisons immédiatement

$$(aX, aY) = |a|^2(X, Y).$$

En particulier, soient un vecteur X de coordonnées $x_1, x_2, \ldots, x_n$, et e_i

le $i^{\text{ième}}$ vecteur coordonné, dont la $i^{\text{ième}}$ composante est égale à 1 et les autres sont nulles; nous avons

$$x_i = (e_i, X).$$

L'espace métrique ainsi défini par l'espace vectoriel et la forme d'Hermite unité s'appellera *espace unitaire*.

2. Nous allons rechercher les *transformations unitaires* de coordonnées, c'est-à-dire les formules de changement de coordonnées qui font *passer d'un système normal à un autre système normal;* la forme d'Hermite unité doit être invariante d'après la définition même de la longueur d'un vecteur.

Si nous désignons par

$$X = S\Xi$$

les formules de transformation de coordonnées, nous devons avoir, quel que soit le vecteur X,

$$\overline{X'}X = \overline{\Xi'}\Xi, \qquad \text{puisque} \quad (X, X) = X'X,$$

c'est-à-dire, *quel que soit le vecteur* Ξ,

$$\overline{\Xi'}\,\overline{S'}S\Xi = \overline{\Xi'}\,\Xi,$$

soit

$$\overline{\Xi'}\,\overline{S'}\,S = \overline{\Xi'},$$

et finalement

$$\overline{S'}S = 1,$$

qui peut encore s'écrire

$$S^*S = 1.$$

Les changements de coordonnées faisant passer d'un système normal à un autre système normal sont obtenus *par une matrice* S *satisfaisant à la relation*

$$\underline{S^*S = 1.}$$

Et si nous nous reportons aux indications déjà données sur la double interprétation des matrices, nous voyons que *les transformations du groupe unitaire, ou transformations unitaires, sont obtenues aussi par une telle matrice* S.

3. Donnons deux propriétés immédiates des *matrices unitaires*. De la relation

$$S^*S = 1,$$

nous déduisons

$$(\text{déterminant de } S^*)(\text{déterminant de } S) = 1;$$

or, pour une matrice quelconque, le déterminant de cette matrice et le déterminant de la matrice associée sont conjugués; il en résulte que

$$|\text{déterminant de } S|^2 = 1,$$

autrement dit une matrice unitaire n'est jamais dégénérée et a *un déterminant dont le module est un.*

D'autre part, la relation de définition des matrices unitaires peut s'écrire

$$S^* = S^{-1};$$

et puisque nous savons qu'une matrice ne peut présenter qu'une seule inverse, nous voyons que la relation de définition peut encore se mettre sous la forme

$$\underline{SS^* = 1,}$$

Dans ce qui précède, on a interprété $X = S\Xi$ comme un changement de coordonnées conservant le caractère normal du système de coordonnées. On peut aussi l'envisager comme une *transformation de l'espace en lui-même* (opérateur), définie par la matrice unitaire S *et conservant les longueurs.* Un tel opérateur sera dit *opérateur unitaire.*

Montrons que les transformations unitaires forment un groupe. D'une part, il est clair que la matrice unité (transformation identique) est unitaire. D'autre part, le produit de deux matrices (ou transformations) unitaires est une matrice (ou transformation) unitaire; en effet, par hypothèse,

$$UU^* = 1,$$
$$VV^* = 1;$$

formons $(UV)(UV)^*$, il vient

$$UVV^*U^* = U(VV^*)U^* = UU^* = 1.$$

Nous pouvons donner de ce résultat une démonstration géométrique très simple : si deux transformations n'altèrent pas la forme fondamentale d'Hermite, leur produit (c'est-à-dire la succession de ces deux transformations) ne l'altère pas non plus, et est, par définition même, une transformation unitaire.

Détaillons algébriquement la relation

$$U^*U = 1,$$

définissant les matrices unitaires; désignant toujours par u_{ik} l'élément de la $i^{\text{ième}}$ ligne et de la $k^{\text{ième}}$ colonne dans la matrice U, il vient

$$\sum_{r=1}^{r=n} \overline{u_{ri}}\, u_{rk} = \delta_{ik} \qquad \text{pour} \quad i \text{ et } k = 1, 2, \ldots, n,$$

c'est-à-dire un ensemble de n^2 relations, analogues aux relations entre les cosinus directeurs d'axes rectangulaires dans la géométrie analytique cartésienne. De même que dans ce cas, elles entraînent n^2 autres relations qui sont

$$\sum_{r=1}^{r=n} u_{ir}\, \overline{u_{kr}} = \delta_{ik} \qquad \text{pour} \quad i \text{ et } k = 1, 2, \ldots, n,$$

que l'on obtient immédiatement en détaillant l'égalité

$$UU^* = 1.$$

Ces relations peuvent s'interpréter géométriquement. Par la relation $X = U\Xi$ nous passons d'un système normal de coordonnées à un autre système normal; désignons par $e_1, e_2, \ldots, e_n$ les vecteurs coordonnés du premier système et par $e'_1, e'_2, \ldots, e'_n$ ceux du second; les quantités $u_{1i}, u_{2i}, \ldots, u_{ni}$ sont les *composantes du vecteur e'_i dans le premier système de coordonnées*; ce système étant normal, la relation

$$\sum_{r} \overline{u_{ri}}\, u_{rk} = \delta_{ik}$$

n'est que le développement du produit scalaire

$$(e'_i, e'_k) = \delta_{ik}.$$

En particulier la longueur des nouveaux vecteurs coordonnés est égale à 1 et le produit scalaire de deux vecteurs coordonnés différents est nul; nous dirons, par définition, que deux tels vecteurs sont orthogonaux. Les n vecteurs e'_i d'un système normal quelconque sont donc n vecteurs unitaires, deux à deux orthogonaux.

4. Nous dirons que deux vecteurs X et Y sont *orthogonaux* si leur produit scalaire est nul :

$$(X, Y) = 0,$$

ou, ce qui revient au même,

$$(Y, X) = 0.$$

Remarquons que n vecteurs unitaires et orthogonaux sont *linéairement indépendants*, car, d'après le paragraphe précédent, le déterminant formé à l'aide de leurs composantes n'est pas nul puisqu'il a 1 pour module.

Démontrons que l'orthogonalité de deux vecteurs est une propriété indépendante du système normal de coordonnées choisi, ou, plus généralement, *que le produit scalaire de deux vecteurs est une grandeur intrinsèque liée aux deux vecteurs*, c'est-à-dire *indépendante du système normal de coordonnées choisi*. En effet, la transformation $X = U\Xi$ respecte la forme fondamentale; par conséquent si nous posons $Y = UH$, et si nous désignons par λ une quantité complexe quelconque, nous avons

$$|X + \lambda Y|^2 = |\Xi + \lambda H|^2,$$

puisque

$$X + \lambda Y = U(\Xi + \lambda A);$$

soit, en développant,

$$|X + \lambda Y|^2 = (X + \lambda Y, X + \lambda Y) = (X, X) + \bar{\lambda}(Y, X) + \lambda(X, Y) + \lambda\bar{\lambda}(Y, Y),$$

et de même

$$|\Xi + \lambda H|^2 = (\Xi + \lambda H, \Xi + \lambda H) = (\Xi, \Xi) + \bar{\lambda}(H, \Xi) + \lambda(\Xi, H) + \lambda\bar{\lambda}(H, H),$$

et, en tenant compte des deux relations

$$(X, X) = (\Xi, \Xi),$$
$$(Y, Y) = (H, H),$$

la nouvelle forme de la relation

$$\bar{\lambda}(Y, X) + \lambda(X, Y) = \bar{\lambda}(H, \Xi) + \lambda(\Xi, H),$$

qui doit être vérifiée quel que soit le nombre complexe λ; cela entraîne

$$(X, Y) = (\Xi, H).$$

Le produit scalaire de deux vecteurs reste donc bien invariant dans un changement unitaire de coordonnées.

Faisons ici une remarque très simple qui nous sera utile par la suite :

Si un vecteur A *est orthogonal à tous les vecteurs* X *de* E_n, *il est nul;* en effet, dans un système normal de coordonnées, nous devons avoir, quelles que soient les composantes x_i,

$$\sum_{i=1}^{n} \bar{a}_i x_i = 0,$$

ce qui entraîne

$$a_i = 0 \quad \text{pour} \quad i = 1, 2, \ldots, n.$$

5. Considérons *l'espace dual* $\mathcal{E}$ de l'espace unitaire E. Si X est un vecteur arbitraire de l'espace E et Ξ un vecteur arbitraire de l'espace $\mathcal{E}$, le produit $\Xi' X$ est invariant dans un changement de coordonnées; or, dans l'espace unitaire, $\overline{X}' X$ est invariant dans un changement de coordonnées; le vecteur $\Xi = \overline{X}$ appartient donc à l'espace dual, c'est-à-dire que nous avons

$$\mathcal{E} = \overline{E}.$$

Nous avons vu que si un opérateur A transforme un espace E_m à m dimensions en un espace E_p à p dimensions, l'opérateur A′ transforme $\mathcal{E}_p$ en $\mathcal{E}_m$, si nous désignons par $\mathcal{E}_m$ l'espace dual de E_m et par $\mathcal{E}_p$ celui de E_h. Or ici

$$\mathcal{E}_m = \overline{E_m},$$
$$\mathcal{E}_p = \overline{E_p}.$$

L'opérateur A′ transforme donc $\overline{E_p}$ en $\overline{E_m}$ et par conséquent l'opérateur $\overline{A'} = A^*$ transforme E_p en E_m. Il en résulte que l'opérateur AA^* transforme E_p en E_m puis E_p, soit finalement E_p en E_p, et que l'opérateur A^*A transforme E_m en E_p puis E_m, soit finalement E_m en E_m.

6. Deux figures composées de vecteurs sont égales dans l'espace unitaire si elles se déduisent l'une de l'autre par une transformation (opérateur) unitaire. *La géométrie unitaire est l'étude des propriétés des figures invariantes dans le groupe unitaire.*

7. Nous allons maintenant *décomposer l'espace en espaces orthogonaux complémentaires.*

Considérons, dans l'espace unitaire, un sous-espace E′ défini par les vecteurs $A_1, A_2, \ldots, A_m$ linéairement indépendants. Puis considérons l'ensemble des vecteurs orthogonaux à tous les A_i, c'est-à-dire orthogonaux à tous les vecteurs de E′; ils forment une multiplicité linéaire E″ satisfaisant aux m équations indépendantes suivantes :

$$(A_1, X) = 0,$$
$$\ldots\ldots\ldots\ldots,$$
$$(A_m, X) = 0;$$

cette multiplicité E″, dont les vecteurs dépendent de $n - m$ constantes

arbitraires, est donc à $n - m$ dimensions. Nous venons, en résumé, de définir deux sous-espaces E′ et E″ tels *qu'un vecteur quelconque de l'un est orthogonal à un vecteur quelconque de l'autre;* nous dirons que ces *deux espaces sont orthogonaux.* Deux tels sous-espaces *ne peuvent avoir de vecteur commun non nul*, car si tel vecteur existait, il devrait être orthogonal à lui-même

$$(X, X) = 0 \qquad \text{soit} \qquad |X|^2 = 0;$$

et nous avons déjà remarqué que cette relation entraîne $X = 0$. Les sous-espaces E′ et E″ de l'espace E_n ayant un nombre total de dimensions égal à $m + (n - m) = n$ et n'ayant pas de vecteur commum, sont nécessairement *complémentaires*

$$E = E' + E''.$$

Cette décomposition jouit d'une propriété très importante : Si un opérateur *unitaire* A *laisse* E′ *invariant il laisse également* E″ *invariant;* car il respecte l'orthogonalité et par conséquent le transformé d'un vecteur orthogonal à E′ est orthogonal à E′, autrement dit il appartient à E″. Donc, *dans le domaine des opérateurs unitaires*, tout sous-espace invariant E′ est accompagné univoquement d'un sous-espace complémentaire orthogonal E″, lui aussi invariant; E″ est une représentation concrète de *l'espace projection de* E *parallèlement à* E′.

8. **Opérateurs hermitiens.** — Soit, dans l'espace unitaire, un opérateur linéaire représenté par la matrice

$$A = \| a_{ik} \|.$$

Nous avons vu dans l'étude de l'espace dual que le produit $\Xi' AY$, dans lequel Ξ est un vecteur arbitraire de l'espace $\mathcal{E}$ dual de E et Y un vecteur arbitraire de E, est invariant par un changement de coordonnées. Or nous savons que dans l'espace unitaire nous pouvons prendre

$$\Xi = \bar{X}.$$

Il en résulte que la forme bilinéaire associée à A,

$$\underline{\bar{X}' AY = (X, AY) = A(X, Y)},$$

est *invariante par toute transformation unitaire.*

Un cas fort intéressant pour les matrices est celui où

$$\boxed{A(Y, X) = \overline{A(X, Y)},}$$

ou, ce qui revient au même,

$$\boxed{(Y, AX) = \overline{(X, AY)} = (AY, X),}$$

quels que soient les vecteurs X *et* Y; nous dirons qu'une telle ***matrice*** est ***hermitienne*** et qu'elle représente un ***opérateur hermitien***. Cette condition peut s'écrire

$$\sum_{i,k} a_{ik}\overline{x}_i y_k = \sum_{i,k} \overline{a}_{ik} y_i \overline{x}_k,$$

soit, en changeant le nom des indices au deuxième membre,

$$\sum_{i,k} a_{ik}\overline{x}_i y_k = \sum_{i,k} \overline{a_{ki}}\,\overline{x}_i y_k.$$

La condition nécessaire et suffisante pour que cette relation soit vérifiée quels que soient x_i et y_i est que, pour toutes valeurs de i et k nous ayons

$$\boxed{a_{ik} = \overline{a_{ki}},}$$

c'est-à-dire

$$\boxed{A = A^{\star}.}$$

Nous retrouvons une définition déjà donnée des matrices hermitiennes.

Mais la définition précédente

$$\boxed{(X, AY) = (AX, Y),}$$

pour tout couple X, Y nous sera aussi très utile dans la suite.

Dans le cas où $Y = X$, la relation fondamentale devient

$$A(X, X) = \overline{A(X, X)},$$

ce qui revient à dire que la forme

$$A(X, X) = \Sigma\, a_{ik}\overline{x}_i x_k$$

est toujours réelle quels que soient les x_i. C'est la ***forme d'Hermite associée à l'opérateur hermitien*** A; ces formes, introduites en algèbre et en arithmétique par Hermite, furent étudiées par Hermite, Jordan, Poincaré et M. Picard. Nous dirons qu'une forme d'Hermite est ***définie positive*** si elle prend une valeur positive pour tout vecteur X non nul;

il résulte de la définition même qu'une *forme définie positive ne peut correspondre à une matrice dégénérée*, car, sinon,

$$A(X, X) = X^* AX$$

s'annulerait pour des vecteurs X non nuls, puisqu'il en serait ainsi de AX. En particulier la *forme d'Hermite unité* $\sum_i \overline{x_i} x_i$ associée à *l'opérateur unité* (qui est un opérateur hermitien) est *définie positive*.

La forme d'Hermite unité définit la longueur d'un vecteur rapporté à un système normal de coordonnées ; si nous avons, au contraire, un système *quelconque* de coordonnées défini par les vecteurs $A_1, A_2, \ldots, A_n$ linéairément indépendants, l'expression de la longueur d'un vecteur sera moins simple. Soit X un vecteur quelconque

$$X = x_1 A_1 + \ldots + x_n A_n.$$

Et alors

$$|X|^2 = (X, X) = (x_1 A_1 + \ldots + x_n A_n, \quad x_1 A_1 + \ldots + x_n A_n) = \Sigma g_{ik} \overline{x_i} x_k,$$

en posant

$$g_{ik} = (A_i, A_k).$$

La forme bilinéaire $\sum_{i,k} g_{ik} \overline{x_i} x_k$ est une forme d'Hermite, car nous avons la relation $g_{ik} = \overline{g_{ki}}$; d'autre part, définissant le carré de la longueur d'un vecteur, elle sera *définie positive*.

Nous allons montrer que, *réciproquement, une forme d'Hermite définie positive, quelconque, peut être considérée comme définissant la longueur des vecteurs dans un système de coordonnées convenablement choisi*. Soit en effet

$$G(X, X) = \sum_{i,k} g_{ik} \overline{x_i} x_k$$

une forme d'Hermite définie positive et la forme bilinéaire associée

$$G(X, Y) = \sum_{i,k} g_{ik} \overline{x_i} y_k ;$$

nous allons construire un *système normal* de coordonnées défini par les vecteurs $e_1, e_2, \ldots, e_n$, tel que dans ce nouveau système nous ayons

$$G(X, X) = \sum_{i=1}^{n} \overline{\xi_i} \xi_i \qquad \text{pour} \quad X = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n.$$

soit encore

$$G[\xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n, \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n] = \Sigma \bar{\xi}_i \xi_i.$$

Or, d'après sa définition, $G(X, Y)$ est linéaire par rapport à X et par rapport à Y, c'est-à-dire que

$$G(X' + X'', Y) = G(X', Y) + G(X'', Y)$$

et de même

$$G(X, Y' + Y'') = G(X, Y') + G(X, Y''),$$

et

$$G(aX, Y) = \bar{a} G(X, Y)$$

et de même

$$G(X, aY) = a G(X, Y).$$

La relation envisagée peut donc s'écrire

$$\sum_{i,k} \bar{\xi}_i \xi_k G(e_i, e_k) = \sum_i \bar{\xi}_i \xi_i.$$

Elle sera satisfaite pour tous les vecteurs X, sous la condition

$$\underline{G(e_i, e_k) = \delta_{ik}} \qquad \underline{\text{pour} \quad i, k = 1, 2, \ldots, n.}$$

Montrons que nous pouvons construire le système de vecteurs e_1, $e_2, \ldots, e_n$ satisfaisant à ces conditions :

Choisissons arbitrairement un vecteur ε_1 non nul; en le multipliant par un nombre réel nous pourrons en déduire *un vecteur* e_1 satisfaisant à $G(e_1, e_1) = 1$ puisque $G(\varepsilon_1, \varepsilon_1) > 0$.

Ceci étant, considérons l'équation $G(e_1, X) = 0$; si n est supérieur à 1, elle admet certainement une solution non nulle au moins; soit ε_2 une telle solution; de même que pour e_1, nous pouvons trouver un nombre réel qui, multipliant ε_2, donne *un vecteur* e_2 satisfaisant à $G(e_2, e_2) = 1$. Nous aurons finalement

$$G(e_i, e_k) = \delta_{ik} \qquad \text{pour} \quad i, k = 1, 2.$$

e_2 est indépendant de e_1 car $G(e_1, e_1) = 1$ et $G(e_1, e_2) = 0$.

Si n est supérieur à 2, le système de deux équations linéaires (indépendantes puisque e_1 et e_2 le sont)

$$G(e_1, X) = 0,$$
$$G(e_2, X) = 0$$

admet au moins une solution; soit ε_3 une telle solution; nous pouvons trouver un nombre réel qui, multipliant ε_3, donne un *vecteur* e_3 satisfai-

sant à $G(e_2, e_3) = 1$. Nous aurons finalement

$$G(e_i, e_k) = \delta_{ik} \quad \text{pour} \quad i, k = 1, 2, 3.$$

Et nous pouvons continuer.

Lorsque nous aurons obtenu *$m < n$ vecteurs indépendants $e_1, e_2, \ldots, e_m$* satisfaisant à

$$G(e_i, e_k) = \delta_{ik} \quad \text{pour} \quad i, k = 1, 2, \ldots, m,$$

nous envisagerons le système suivant de m équations linéaires (indépendants comme les $e_1, \ldots, e_m$) :

$$G(e_i, X) = 0 \quad \text{pour} \quad i = 1, 2, \ldots, m;$$

il admet au moins une solution non nulle; soit e_{m+1} une telle solution; nous pouvons alors trouver un nombre réel qui, multipliant ε_{m+1}, donne *un vecteur* e_{m+1} satisfaisant à $G(e_{m+1}, e_{m+1}) = 1$. L'ensemble de $m+1$ vecteurs ainsi formés satisfera à

$$G(e_i, e_k) = \delta_{ik} \quad \text{pour} \quad i, k = 1, 2, \ldots, m, m+1.$$

Nous obtiendrons finalement n vecteurs $e_1, e_2, \ldots, e_n$ satisfaisant aux conditions requises. Mais il faut nous assurer qu'ils peuvent être pris comme vecteurs coordonnés, c'est-à-dire qu'ils sont linéairement indépendants. Soit à démontrer que la relation

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n = 0$$

ne peut être satisfaite que si tous les λ sont nuls; en effet nous aurions

$$G[\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n, \lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n] = 0,$$

soit, en développant,

$$\sum_{i,k} \overline{\lambda_i} \lambda_k G(e_i, e_k) = 0,$$

et, en tenant compte des relations liant les e_i,

$$\sum_{i=1}^{n} \overline{\lambda_i} \lambda_i = 0$$

ou

$$\Sigma \mid \lambda_i \mid^2 = 0,$$

qui exige la nullité de tous les λ_i.

L'ensemble des e_i définit donc un système de coordonnées; nous pouvons dire que ce *système de coordonnées est normal* et que *la forme*

d'Hermite G(X, X) *définit la longueur du vecteur* X. Cette réciproque a des *applications* intéressantes.

1° Nous avons, dans ce qui précède, introduit la métrique par une expression analytique de la longueur du vecteur, en fonction des composantes de ce vecteur dans un système normal de coordonnées.

Axiomatiquement, on développe directement une *métrique de l'espace affine* lorsqu'on stipule que :

A tout vecteur X on fait correspondre un nombre réel représenté par $|X|^2 = (X, X)$, que l'on appelle carré de sa longueur, et qui est une *forme d'Hermite définie positive du vecteur* X.

Il est clair que cette façon de procéder est entièrement équivalente à celle employée au début de ce chapitre, puisque nous venons de voir qu'il est toujours possible de se ramener à un système normal de coordonnées.

2° Le procédé précédent a une *autre application* très importante : il permet *d'adapter un système normal de coordonnées à une décomposition de l'espace;* soit en effet E′ un sous-espace à m dimensions de l'espace E à n dimensions.

Nous venons de démontrer qu'il est toujours possible de choisir dans E′, m vecteurs normaux linéairement indépendants et deux à deux orthogonaux; poursuivant la méthode dans l'espace E, nous associerons à ces m vecteurs, $n - m$ autres vecteurs *orthonormaux* (¹). L'ensemble des vecteurs $e_1, e_2, \ldots, e_n$ détermine pour E un *système normal de coordonnées adapté à la décomposition de* E *en deux sous-espaces* E′ *et* E″; l'ensemble des vecteurs $e_1, e_2, \ldots, e_m$ détermine pour E′ un système normal de coordonnées tandis que l'ensemble des vecteurs e_{m+1}, $e_{m+2}, \ldots, e_n$ détermine pour E″ un système normal de coordonnées. Ces deux sous-espaces sont tels, par leur définition, qu'un vecteur quelconque de l'un est orthogonal à un vecteur quelconque de l'autre; nous dirons qu'ils sont *orthogonaux*.

9. Cherchons si les opérateurs hermitiens forment un groupe; il faudrait pour cela que, si A et B sont hermitiens, leur produit AB le soit, c'est-à-dire que

$$BA = (AB)^* = B^* A^*,$$

(¹) Un système de vecteurs *orthonormaux* est formé de vecteurs *normaux* (longueur unité), deux à deux *orthogonaux*.

soit, en tenant compte des hypothèses $A = A^*$, $B = B^*$,

$$AB = BA.$$

Par conséquent, *le produit de deux opérateurs hermitiens n'est hermitien que si ces deux opérateurs sont permutables.* Les opérateurs hermitiens ne forment pas un groupe.

Les éléments d'une matrice hermitienne, situés dans la diagonale principale sont *réels;* il en est donc de même de leur somme qui est la *trace de la matrice.*

Considérons un opérateur quelconque A; nous lui associerons les deux opérateurs AA^* et A^*A que nous appellerons ses *normes hermitiennes;* ces deux opérateurs sont, en effet, hermitiens

$$(AA^*)^* = A^{**}A^* = AA^*,$$

et de même

$$(A^*A)^* = A^*A^{**} = A^*A.$$

Ces deux matrices hermitiennes ont même trace [1] $\sum_{i,k} |a_{ik}|^2$, que nous appellerons carré de la *valeur absolue de la matrice* A.

Pour terminer, signalons deux relations très importantes :

$$AA^*(X, X) = \sum_{i,k} \Big(\sum_l a_{il}\bar{a}_{kl} \Big) \bar{x}_i x_k = \sum_l \Big(\sum_i a_{il}\bar{x}_i \Big) \Big(\sum_k \bar{a}_{kl} x_k \Big),$$

c'est-à-dire

$$AA^*(X, X) = \sum_{l=1}^{n} \Big| \sum_{i=1}^{n} a_{il}\bar{x}_i \Big|^2$$

et de même

$$A^*A(X, X) = \sum_{l=1}^{n} \Big| \sum_{i=1}^{n} a_{li} x_i \Big|^2.$$

10. Tout vecteur X non nul détermine un *rayon*, en désignant ainsi l'ensemble des vecteurs λX, où λ est un nombre complexe arbitraire. Les vecteurs normés du rayon correspondent à une valeur de λ satisfaisant à

$$|\lambda| . |X| = 1,$$

c'est-à-dire

$$|\lambda| = \frac{1}{|X|};$$

---

[1] *Voir* au Chapitre précédent la méthode de Frobenius pour la métrisation de l'espace des matrices.

nous voyons que λ n'est déterminé qu'en *module;* en particulier, nous dirons que le vecteur $\frac{X}{|X|}$ est le *vecteur normé porté par* X; l'ensemble des vecteurs normés du rayon est donné par la formule

$$\frac{X}{|X|}e^{i\theta},$$

où θ désigne un angle réel arbitraire.

Toute transformation A de l'espace, à déterminant non nul, est aussi une transformation du corps des rayons; mais, *dans le corps des rayons*, les deux transformations A et λA (où λ désigne un nombre complexe arbitraire, non nul) ont le même effet. Toute transformation A qu'on peut, après multiplication par un scalaire, ramener à une transformation unitaire, sera dite une *rotation* dans le corps des rayons. Deux transformations unitaires S et S', telles que

$$S' = \varepsilon S,$$

ε désignant un nombre complexe arbitraire de module 1, sont des *rotations identiques* dans le corps des rayons.

II. — QUELQUES APPLICATIONS IMMÉDIATES ET UTILES DE CES NOTIONS DE MÉTRIQUE.

1. Inégalité de Cauchy-Schwarz. — Si X et Y sont deux vecteurs quelconques, leur produit scalaire et leurs modules satisfont à l'inégalité

$$|(X, Y)| \leqq |X|.|Y|.$$

Pour simplifier la démonstration, nous allons nous placer dans un système de coordonnées normales; désignons par $x_1, x_2, \ldots, x_n$ les coordonnées de X dans ce système, et par $y_1, y_2, \ldots, y_n$ celles de Y; l'inégalité s'écrit alors

$$\left|\sum_{i=1}^{n}\bar{x}_i y_i\right|^2 \leqq \left[\sum_{i=1}^{n}\bar{x}_i x_i\right] \times \left[\sum_{i=1}^{n}\bar{y}_i y_i\right],$$

soit encore

$$\left|\sum_{i=1}^{n}\bar{x}_i y_i\right|^2 \leqq \left[\sum_{i=1}^{n}|x_i|^2\right].\left[\sum_{i=1}^{n}|y_i|^2\right].$$

Cette inégalité se déduit immédiatement de la superposition des deux

suivantes :

$$\left|\sum_{i=1}^{n}\bar{x}_i y_i\right|^2 \leqq \left[\sum_{i=1}^{n}|x_i||y_i|\right]^2,$$

$$\left[\sum_{i=1}^{n}|x_i|.|y_i|\right]^2 \leqq \left[\sum_{i=1}^{n}|x_i|^2\right]\times\left[\sum_{i=1}^{n}|y_i|^2\right].$$

La première de ces inégalités exprime simplement que le module d'une somme de nombres complexes est inférieur à la somme des modules de ces nombres. La seconde est une inégalité très importante due à Cauchy et utilisée par Schwarz (qui l'a généralisée en remplaçant les sommes par des intégrales) dans l'étude du problème de Dirichlet; sa démonstration est immédiate :

$$\sum_{i=1}^{n}[\lambda|x_i|+\mu|y_i|]^2=\lambda^2\sum_{i=1}^{n}|x_i|^2+2\lambda\mu\sum_{i=1}^{n}|x_i||y_i|+\mu^2\sum_{i=1}^{n}|y_i|^2;$$

or le premier membre ne peut devenir négatif; il en est donc de même de la forme quadratique située au second membre; cette forme a donc un discriminant négatif ou nul, soit

$$[\Sigma|x_i||y_i|]^2-[\Sigma|x_i|^2].[\Sigma|y_i|^2]\leqq 0,$$

qui n'est autre que l'inégalité cherchée.

Nous ne pourrons avoir

$$|(X, Y)| = |X|.|Y|$$

que si les deux inégalités intermédiaires deviennent des égalités; pour la seconde, cela exige que la forme quadratique puisse s'annuler, c'est-à-dire qu'il existe des valeurs convenables de λ et μ telles que

$$\lambda|x_i|+\mu|y_i|=0 \qquad \text{pour} \qquad i=1, 2, \ldots, n;$$

ce qui peut s'exprimer de la façon suivante : *les rapports* $\frac{y_i}{x_i}$ *ont un module indépendant de l'indice* i. La premiére inégalité intermédiaire deviendra une égalité à la condition que les quantités $\bar{x}_i y_i$ aient toutes *même argument* ou bien encore que les rapports $\frac{y_i}{x_i}$ aient un argument indépendant de l'indice i. Finalement nous voyons qu'*il y aura égalité dans le cas, et seulement dans le cas,* où

$$\frac{y_i}{x_i}=\lambda,$$

c'est-à-dire

$$Y = \lambda X,$$

λ étant un nombre complexe arbitraire.

Nous pouvons énoncer ce théorème sous la forme suivante :

Si A *est un vecteur quelconque et* X *un vecteur de longueur* 1, *on a toujours*

$$|(A, X)| \leqq |A|,$$

l'égalité n'ayant lieu que pour

$$X = \frac{A}{|A|} e^{i\theta},$$

relation où θ *désigne un angle réel arbitraire.* Dans le cas particulier où A est nul, l'égalité est satisfaite pour tous les vecteurs X.

2. Systèmes de vecteurs orthogonaux. Méthode d'orthogonalisation de Schmidt. — Un système de vecteurs est dit *orthogonal*, si les vecteurs qui le composent sont deux à deux orthogonaux. Il est dit orthogonal et normal, ou plus brièvement *orthonormal*, si, de plus, les vecteurs ont une longueur égale à 1; si $e_1, e_2, \ldots, e_m$ désignent les vecteurs du système, ces conditions peuvent se résumer par

$$\boxed{(e_i, e_k) = \delta_{ik}.}$$

Les vecteurs formant un système orthonormal sont nécessairement *linéairement indépendants;* montrons en effet que la relation

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_m e_m = 0$$

ne peut avoir lieu que si tous les λ sont nuls; il suffit, pour cela, de multiplier scalairement par e_i :

$$\lambda_1(e_1, e_i) + \lambda_2(e_2, e_i) + \ldots + \lambda_i(e_i, e_i) + \ldots + \lambda_m(e_m, e_i) = 0;$$

ce qui, compte tenu des hypothèses, se réduit à

$$\lambda_i = 0.$$

Un jeu de m vecteurs orthonormaux forme donc un ensemble de m vecteurs linéairement indépendants. En particulier, avec un jeu de n vecteurs orthonormaux de l'espace vectoriel à n dimensions, on peut *exprimer tous les vecteurs de cet espace*, car les vecteurs $X, e_1, e_2, \ldots, e_n$

au nombre de $n+1$ sont certainement dépendants, soit

$$\lambda X + \lambda_1 e_1 + .. + \lambda_n e_n = 0.$$

Or λ n'est certainement pas nul, puisque les vecteurs $e_1, e_2, \dots, e_n$ sont indépendants; nous pouvons donc résoudre cette relation par rapport à X :

$$X = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$$

Les vecteurs $e_1, e_2, \dots, e_n$ *peuvent donc servir de base à l'espace* E_n. Les nombres x_i sont appelés les *composantes* du vecteur X par rapport au système de coordonnées défini par les vecteurs e_i; ils ont une expression particulièrement simple, obtenue en multipliant scalairement à gauche la dernière relation par e_i :

$$x_i = (e_i, X).$$

Arrivons maintenant au *procédé d'orthogonalisation de Schmidt :*

Soient m vecteurs linéairement indépendants $A_1, A_2, \dots, A_m$ $(m \leqq n)$. *Ces vecteurs déterminent une multiplicité linéaire; nous pouvons trouver dans cette multiplicité m vecteurs orthogonaux et normaux* (et par conséquent linéairement indépendants), *définis par des relations de la forme*

$$\begin{aligned} B_1 &= r_{11} A_1, \\ B_2 &= r_{21} A_1 + r_{22} A_2, \\ B_3 &= r_{31} A_1 + r_{32} A_2 + r_{33} A_3, \\ &\dots\dots\dots\dots\dots\dots, \\ B_m &= r_{m1} A_1 + r_{m2} A_2 + r_{m3} A_3 + \dots + r_{mm} A_m, \end{aligned}$$

les r_{ik} *étant choisis tels que*

$$(B_i B_k) = \delta_{ik}$$

et $r_{ii} \neq 0$ *pour toute valeur de i, de telle sorte qu'inversement* A_1 *s'exprime en fonction de* B_1, A_2 *en fonction de* B_1 *et de* B_2, A_3 *en fonction de* B_1, B_2 *et* B_3, *etc.*

Une matrice telle que $\| r_{ik} \|$, dont tous les éléments situés au-dessus de la diagonale principale sont nuls, ceux de la diagonale principale ne l'étant pas, est dite *matrice de récurrence* ou encore *matrice canonique non dégénérée* [1].

---

[1] La matrice est canonique en ce sens que tous les éléments situés au-dessus de la diagonale principale sont nuls; elle est non dégénérée, car son déterminant, qui est égal à $r_{11} r_{22} \dots r_{mm}$, est différent de zéro.

La détermination des vecteurs orthonormaux sera faite en deux étapes; dans la première nous allons former un système de vecteurs orthogonaux, dans la seconde nous normerons.

Première étape. — Cherchons un système de vecteurs orthogonaux $\beta_1, \beta_2, \ldots, \beta_m$ définis par des relations de la forme

$$\begin{aligned}
\beta_1 &= A_1,\\
\beta_2 &= \rho_{21}\beta_1 + A_2,\\
\beta_3 &= \rho_{31}\beta_1 + \rho_{32}\beta_2 + A_3,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\
\beta_m &= \rho_{m1}\beta_1 + \rho_{m2}\beta_2 + \rho_{m3}\beta_3 + \ldots + A_m.
\end{aligned}$$

Nous allons déterminer de proche en proche les coefficients ρ. ρ_{21} sera obtenu en écrivant que les vecteurs β_1 et β_2 sont orthogonaux

$$(\beta_1, \beta_2) = (A_1, A_2) + \rho_{21}(A_1, \beta_1) = 0;$$

et puisque $(A_1, \beta_1) = |A_1|^2$ n'est pas nul, nous obtenons

$$\rho_{21} = -\frac{(A_1, A_2)}{|A_1|^2},$$

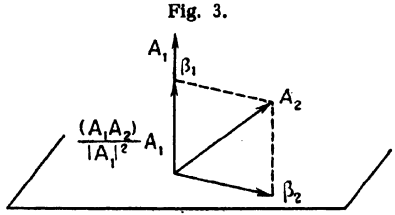
et le vecteur β_2 non nul :

$$\beta_2 = A_2 - \frac{(A_1, A_2)}{|A_1|^2} A_1.$$

($\beta_2 \neq 0$, car A_1 et A_2 sont *indépendants*). Cette formule a une interprétation géométrique très simple

On obtient β_2 *en retranchant de* A_2 *sa composante suivant le vecteur* A_1.

Fig. 3.



Nous allons voir que, de même, nous obtiendrons β_3 en retranchant de A_3 sa composante dans le plan déterminé par les deux vecteurs A_1 et A_2. En effet :

β_3 doit être orthogonal à β_1 et β_2, ce qui nous conduit aux deux

relations :

$$(\beta_1, \beta_3) = (\beta_1, A_3) + \rho_{32}(\beta_1, \beta_2) + \rho_{31}|\beta_1|^2 = 0,$$
$$(\beta_2, \beta_3) = (\beta_2, A_3) + \rho_{32}|\beta_2|^2 \quad + \rho_{31}(\beta_2 \beta_1) = 0,$$

soit

$$(\beta_1, A_3) + \rho_{31}|\beta_1|^2 = 0,$$
$$(\beta_2, A_3) + \rho_{32}|\beta_2|^2 = 0;$$

les vecteurs β_1 et β_2 n'étant pas nuls, nous pouvons tirer ρ_{31} et ρ_{32} de ces formules et par suite obtenir le vecteur β_3. Les trois vecteurs ainsi obtenus sont orthogonaux deux à deux; et nous pouvons continuer ainsi de proche en proche, *aucun des vecteurs β_i n'étant nul*, puisque β_2 est une combinaison linéaire à coefficients non tous nuls de A_1 et A_2 linéairement indépendants, β_3 est une combinaison linéaire de A_3, β_2 et β_1, c'est-à-dire de A_3, A_2 et A_1 linéairement indépendants, et ainsi de suite.

Seconde étape. — Nous pouvons alors franchir la *seconde étape* en normant les vecteurs β que nous venons de déterminer. Il suffit de poser

$$B_i = \frac{\beta_i}{|\beta_i|}.$$

Les vecteurs B_i sont deux à deux orthogonaux, puisque les β_i le sont, et les coefficients $r_{ii} = \frac{1}{|\beta_i|}$ sont bien différents de zéro.

Considérons maintenant le cas où les vecteurs donnés $A_1, A_2, \ldots, A_m$ *ne sont pas linéairement indépendants*. Supposons donc que les p premiers vecteurs $A_1, A_2, \ldots, A_p$ soient linéairement indépendants et que le vecteur A_{p+1} en soit une combinaison linéaire. Le *vecteur* β_{p+1}, que nous obtenons en retranchant du vecteur A_{p+1} sa composante dans la multiplicité définie par $A_1, A_2, \ldots, A_p$, *est nul;* il nous suffira de continuer les opérations de Schmidt sans plus nous occuper du vecteur β_{p+1}. Finalement, si nous appliquons brutalement la méthode de Schmidt à des vecteurs qui ne sont pas indépendants, *nous obtiendrons autant de fois un vecteur β nul qu'il y a de vecteurs* A *combinaisons linéaires des vecteurs précédents;* autrement dit, il y aura autant de vecteurs β nuls qu'il y a de relations *distinctes* entre les vecteurs A. Les vecteurs β *non nuls* pourront alors être normés et servir de base à la multiplicité qui a un nombre de dimensions inférieur à m.

3. Systèmes complets. — Nous dirons qu'un système de vecteurs B_i orthogonaux est *complet*, s'il n'existe pas de vecteur X non nul orthogonal à tous les vecteurs du système, c'est-à-dire vérifiant

$$(B_i, X) = 0 \qquad \text{pour} \qquad i = 1, 2, \ldots, m.$$

Le système sera dit *incomplet* s'il existe un vecteur non nul satisfaisant à ces équations.

Si le nombre m des vecteurs orthogonaux est inférieur au nombre de dimensions n de l'espace, le système de vecteurs orthogonaux est *incomplet*, car un système de m équations linéaires homogènes à n inconnues avec $m < n$ présente au moins une solution non nulle.

Si le nombre de vecteurs orthogonaux est égal au nombre de dimensions de l'espace, le système est *complet*, car le système de n équations linéaires homogènes à n inconnues n'a que la solution o, étant donné que son déterminant n'est pas nul (les vecteurs B_i, formant un système orthogonal, sont linéairement indépendants).

Nous avons donc un critère simple, *basé sur le nombre de dimensions*, qui nous permet de reconnaître si un système de vecteurs orthogonaux est complet ou non; mais ce critère ne peut évidemment pas s'étendre aux espaces à une infinité de dimensions; nous allons donner *un autre critère*, *permettant l'extension à une infinité de dimensions*, et que l'on désigne par l'*inégalité de Parseval* ou *de Bessel*.

Soit un système orthogonal et normal de vecteurs $e_1, e_2, \ldots, e_m$ et un vecteur X, arbitraire, de E_n. Ce vecteur est la somme de deux autres : l'un situé dans la multiplicité linéaire définie par le système orthogonal, l'autre dans la multiplité complémentaire (si elle existe)

$$X = X_m + x_1 e_1 + x_2 e_2 + \ldots + x_m e_m,$$

soit encore

$$X_m = X - \sum_{i=1}^{m} x_i e_i.$$

Élevons les deux membres au carré scalaire

$$|X_m|^2 = \left(X - \sum_{i=1}^{m} x_i e_i,\ X - \sum_{i-1}^{m} x_i e_i \right)$$

$$= |X|^2 - \left(X, \sum_{i=1}^{m} x_i e_i \right) - \left(\sum_{i=1}^{m} x_i e_i, X \right) + \left| \sum_{i=1}^{x} x_i e_i \right|^2.$$

Utilisons maintenant l'hypothèse : le système des e_i est orthonormal. Donc

$$(e_i, X) = x_i,$$

$$(X, e_i) = \bar{x}_i,$$

$$(e_i, e_k) = \delta_{ik}.$$

Il vient alors

$$|X_m|^2 = |X|^2 - \sum_{i=1}^{m} \bar{x}_i x_i - \sum_{i=1}^{m} \bar{x}_i x_i + \sum_{i=1}^{m} \bar{x}_i x_i;$$

soit

$$|X_m|^2 = |X|^2 - \sum_{i+1}^{m} \bar{x}_i x_i.$$

Par conséquent nous avons *toujours*, *quel que soit* X.

$$\sum_{i=1}^{m} \bar{x}_i x_i \leqq |X|^2,$$

l'égalité ayant lieu lorsque X_m est nul, et dans ce cas seulement. Or :

Si le système est *complet*, le vecteur X_m, qui doit être orthogonal à $e_1, e_2, \ldots, e_m$, est nécessairement nul quel que soit X, par conséquent, pour un vecteur X *quelconque*, nous avons l'égalité de Pythagore

$$|X|^2 = \sum_{i=1}^{m} \bar{x}_i x_i.$$

Si le système est *incomplet*, *l'égalité ne peut avoir lieu pour tous les vecteurs* X, car il existe au moins un vecteur non nul X orthogonal aux vecteurs $e_1, e_2, \ldots, e_m$, donc tel que

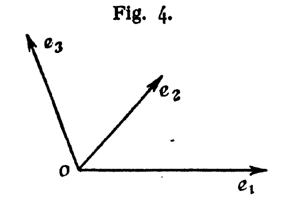
$$X = X_m \neq 0 \qquad \text{et} \qquad x_1 = x_2 = \ldots = x_m = 0.$$

En résumé, *pour que le système soit complet, il faut et il suffit que, pour tout vecteur* X, *on ait*

$$|X|^2 = \sum_{i=1}^{m} \bar{x}_i x_i.$$

Pour un système incomplet, l'égalité précédente n'est réalisée que

Fig. 4.



pour un vecteur X situé dans la multiplicité définie par le système

incomplet de vecteurs. Dans le cas du système complet, les vecteurs $e_1, e_2, \ldots, e_m$ $(m = n)$ déterminent un système de coordonnées normales pour l'espace E_n.

Les résultats auxquels nous venons d'aboutir sont évidents dans le cas de la géométrie euclidienne à trois dimensions.

Trois vecteurs orthogonaux forment un système *complet;* il n'existe pas de vecteur non nul orthogonal à ces trois vecteurs; si x_1, x_2, x_3 désignent les projections orthogonales d'un vecteur X de l'espace sur ces trois vecteurs, nous avons

$$X^2 = x_1^2 + x_2^2 + x_3^2 \qquad \text{(géométrie réelle)}$$

et enfin ces trois vecteurs déterminent un système de coordonnées pour l'espace

$$X = x_1 e_1 + x_2 e_2 + x_3 e_3.$$

Deux vecteurs orthogonaux forment un système *incomplet;* il existe un vecteur orthogonal à ces deux vecteurs; si x_1, x_2 désignent les projections orthogonales d'un vecteur X de l'espace sur ces deux vecteurs, nous avons

$$X^2 \geqq x_1^2 + x_2^2,$$

l'égalité n'étant réalisée que pour les vecteurs X situés dans le plan défini par les vecteurs e_1 et e_2; enfin ces deux vecteurs ne déterminent pas un système de coordonnées de l'espace.

Le critère que nous venons d'obtenir peut revêtir une forme plus générale :

La condition nécessaire et suffisante pour qu'un système orthonormal soit complet est que, quel que soient les vecteurs X *et* Y, *on ait*

$$(X, Y) = \sum_{i=1}^{m} \bar{x}_i y_i.$$

La première forme du critère se déduit de celle-ci en faisant $X = Y$; réciproquement cette relation se déduit de la première appliquée au vecteur $X + \lambda Y$, où λ désigne un scalaire quelconque :

$$|X + \lambda Y|^2 = \sum_{i=1}^{m} |x_i + \lambda y_i|^2$$

ou

$$(X + \lambda Y, X + \lambda Y) = \sum_{i=1}^{m} (\bar{x}_i + \bar{\lambda}\bar{y}_i).(x_i + \lambda y_i).$$

Soit, en développant

$$|X|^2 + \lambda(X, Y) + \bar{\lambda}(Y, X) + \lambda\bar{\lambda}|Y|^2 = \sum_{i=1}^{m} [\bar{x}_i x_i + \lambda \bar{x}_i y_i + \bar{\lambda} \bar{y}_i x_i + \lambda\bar{\lambda} \bar{y}_i y_i].$$

Et en tenant compte des deux relations :

$$|X|^2 = \sum_{i=1}^{m} \bar{x}_i x_i,$$

$$|Y|^2 = \sum_{i=1}^{m} \bar{y}_i y_i,$$

il vient

$$\lambda(X, Y) + \bar{\lambda}(Y, X) = \lambda \sum_{i=1}^{m} \bar{x}_i y_i + \bar{\lambda} \sum_{i=1}^{m} \bar{y}_i x_i.$$

qui doit être vérifiée quel que soit λ complexe; il en résulte

$$(X, Y) = \sum_{i=1}^{m} \bar{x}_i y_i.$$

et, par suite,

$$(Y, X) = \sum_{i=1}^{m} \bar{y}_i x_i.$$

Ainsi que nous l'avons déjà signalé, ce critère, sous l'une quelconque de ses deux formes, est plus intéressant que le critère de la dimension, car *il peut se généraliser à un espace à une infinité de dimensions.*

4. Critère de Gram pour que *m* vecteurs soient linéairement indépendants. — L'intérêt de ce critère réside en ce qu'il ne fait pas intervenir les coordonnées des vecteurs, mais uniquement des éléments *invariants par transformation unitaire*, autrement dit des éléments ayant une *existence intrinsèque* dans la géométrie unitaire; ces éléments seront ici des produits scalaires. Le critère de Gram s'énonce de la façon suivante :

La condition nécessaire et suffisante pour que m vecteurs $A_1, A_2, \ldots, A_m$ *soient linéairement dépendants est que le déterminant*

$$\Gamma = \begin{vmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ A_{m1} & A_{m2} & \ldots & A_{mm} \end{vmatrix}$$

soit nul, A_{ik} *désignant le produit scalaire* (A_i, A_k). *Lorsque les vecteurs* $A_1, A_2, \ldots, A_m$ *sont indépendants, ce déterminant n'est pas nul et il est positif.*

Donnons la démonstration de Schmidt [1] :

Si nous effectuons la transformation

$$A_i = \sum_{k=1}^{m} \gamma_{ik} C_k,$$

successivement dans le premier et le deuxième terme de chaque produit scalaire figurant dans Γ, le déterminant de Gram devient successivement

$$\Gamma = \overline{\begin{vmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \ldots & \gamma_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mm} \end{vmatrix}} \begin{vmatrix} (C_1, A_1) & (C_1, A_2) & \ldots & (C_1, A_m) \\ (C_2, A_1) & (C_2, A_2) & \ldots & (C_2, A_m) \\ \ldots & \ldots & \ldots & \ldots \\ (C_m, A_1) & (C_m, A_2) & \ldots & (C_m, A_m) \end{vmatrix}$$

$$= \overline{|\gamma_{ik}|}\, |\gamma_{ik}|\, |(C_i, C_k)|.$$

En particulier, nous pouvons prendre pour C_i les vecteurs β_i obtenus au n° 2 par la méthode d'orthogonalisation de Schmidt; nous savons que, dans ces conditions,

$$|\gamma_{ik}| = 1$$

et

$$(C_i, C_k) = 0 \qquad \text{si} \quad i \neq k,$$

tandis que $(C_i, C_i) = |\beta_i|^2$; la relation précédente devient alors

$$\Gamma = |\beta_1|^2 . |\beta_2|^2 \ldots |\beta_m|^2.$$

Le déterminant de Gram est donc toujours *positif ou nul;* la condition nécessaire et suffisante pour qu'il soit nul est que l'un des vecteurs β_i soit nul, c'est-à-dire que les vecteurs $A_1, A_2, \ldots, A_m$ soient linéairement indépendants.

Indiquons une autre démonstration, en admettant quelques résultats que nous trouverons dans le paragraphe suivant. Le déterminant Γ se présente dans l'étude de la forme d'Hermite

$$C(X, X) = |x_1 A_1 + x_2 A_2 + \ldots + x_n A_n|^2 = \sum_{i,k} \bar{x}_i x_k (A_i, A_k).$$

Si nous envisageons les systèmes de nombres x_i tels que $\Sigma |x_i|^2 = 1$, la

(1) Voir *Rendi conti del Circolo mathematico di Palermo*, 1908.

forme d'Hermite est définie positive et ne peut s'annuler que si les vecteurs donnés sont linéairement dépendants.

Nous verrons que cela entraîne :

Dans le cas où les vecteurs donnés sont indépendants, que toutes les valeurs du spectre de la matrice $\| A_{ik} \|$ sont positives;

Dans le cas où les vecteurs donnés sont dépendants, que l'une au moins des valeurs du spectre est nulle.

Les valeurs caractéristiques étant données par l'équation

$$| A_{ik} - \lambda \delta_{ik} | = 0,$$

il en résulte que le déterminant Γ de la matrice $\| A_{ik} \|$ est égal au produit des valeurs caractéristiques; donc :

Dans le cas où les vecteurs donnés sont indépendants, Γ est positif;

Dans le cas où les vecteurs donnés sont dépendants, Γ est nul.

Nous pouvons encore raisonner de la façon suivante : prenons un système normal de coordonnées et désignons par $x_{i1}, x_{i2}, \ldots, x_{in}$ les coordonnées du vecteur A_i. Nous aurons alors

$$A_{ik} = \sum_{l=1}^{n} \bar{x}_{il} x_{kl}$$

et le déterminant de Gram pourra s'écrire

$$\Gamma = \left| \sum_{l} \bar{x}_{il} x_{kl} \right| = \Sigma \, | \bar{x}_{il_k} x_{kl_k} |,$$

le dernier signe Σ étant étendu à tous les arrangements possibles de m entiers l_k compris entre 1 et n. Le déterminant correspondant est nul si deux de ces entiers sont égaux; il en résulte que Γ est nul si $m > n$. Or, dans ce cas, les m vecteurs sont nécessairement dépendants.

Il nous reste donc uniquement à examiner le cas $m \leqq n$. La même propriété de décomposition d'un déterminant en une somme de déterminants nous permet aisément de voir que

$$\Sigma \, | \bar{x}_{il_k} x_{kl_k} | = \Sigma' \begin{vmatrix} \bar{x}_{1l_1} x_{1l_1} + \ldots + \bar{x}_{1l_m} x_{1l_m} & \bar{x}_{1l_1} x_{2l_1} + \ldots + \bar{x}_{1l_m} x_{2l_m} & \ldots & \bar{x}_{1l_1} x_{ml_1} + \ldots + \bar{x}_{1l_m} x_{ml_m} \\ \bar{x}_{2l_1} x_{1l_1} + \ldots + \bar{x}_{2l_m} x_{1l_m} & \bar{x}_{2l_1} x_{2l_1} + \ldots + \bar{x}_{2l_m} x_{2l_m} & \ldots & \bar{x}_{2l_1} x_{ml_1} + \ldots + \bar{x}_{2l_m} x_{ml_m} \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots\ldots \\ \bar{x}_{ml_1} x_{1l_1} + \ldots + \bar{x}_{ml_m} x_{1l_m} & \bar{x}_{ml_1} x_{2l_1} + \ldots + \bar{x}_{ml_m} x_{2l_m} & \ldots & \bar{x}_{ml_1} x_{ml_1} + \ldots + \bar{x}_{ml_m} x_{ml_m} \end{vmatrix}$$

le signe Σ' étant étendu à toutes les combinaisons possibles des nombres entiers $l_1, l_2, \ldots, l_m$ tel que

$$1 \leqq l_1 < l_2 < \ldots < l_m \leqq n.$$

Or, au second membre, nous reconnaissons le produit d'une matrice, par la matrice conjuguée; Γ est donc la somme des carrés des modules des déterminants d'ordre m extraits de la matrice

$$\begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{Vmatrix}$$

formée avec les composantes des vecteurs $A_1, A_2, \ldots, A_n$. Nous retrouvons ainsi le critère de Gram.

Si les vecteurs sont indépendants, ces déterminants ne sont pas tous nuls et le déterminant Γ est positif;

Si les vecteurs sont dépendants, ces déterminants sont tous nuls et le déterminant Γ est nul.

III. — OPÉRATEURS ET MATRICES REMARQUABLES DANS L'ESPACE MÉTRIQUE. FORMES CANONIQUES DANS LE GROUPE UNITAIRE.

1. Forme canonique ou réduite de Schur. — Si nous effectuons dans l'espace unitaire, un changement d'axes de coordonnées défini par

$$X = U\Xi,$$

l'opérateur linéaire A qui s'exprimait dans les anciennes coordonnées par la matrice A s'exprime dans les nouvelles coordonnées par la matrice $U^{-1}AU$. Les deux matrices A et $U^{-1}AU$ sont dites équivalentes dans la géométrie unitaire.

Si nous effectuons sur l'espace unitaire une transformation linéaire définie par

$$\mathfrak{X} = UX,$$

l'opérateur A est transformé en l'opérateur UAU^{-1}. Les deux matrices A et UAU^{-1} sont équivalentes dans la géométrie unitaire.

Deux matrices équivalentes à l'un de ces points de vue, le sont également à l'autre; la recherche d'une *forme canonique* d'une matrice dans la géométrie unitaire, consiste à trouver une matrice unitaire U telle que,

par exemple, la matrice $U^{-1}AU$ *ait la forme la plus simple possible*, cette matrice $U^{-1}AU$ sera alors dite la *forme canonique, dans le groupe unitaire*, de la matrice A.

Or, de combien de paramètres dépend la matrice unitaire la plus générale ? Elle comprend n^2 éléments liés par n relations de normalisation et $\frac{n(n-1)}{2}$ relations d'orthogonalité; elle dépend donc finalement de

$$n^2 - n - \frac{n(n-1)}{2} = \frac{n(n-1)}{2}$$

paramètres; d'autre part, une matrice d'ordre n comprend $\frac{n(n-1)}{2}$ éléments situés au-dessus de sa diagonale principale; dans le cas général, nous ne pouvons donc espérer guère mieux que d'*annuler les termes situés au dessus de la diagonale principale dans la forme canonique*.

Nous allons montrer, de proche en proche, qu'il est possible de trouver un système de coordonnées normales $(e_1, e_2, \ldots, e_n)$ tel qu'il en soit ainsi; la forme canonique obtenue est appelée *forme de Schur*. Nous envisageons donc *l'opérateur linéaire* A et nous cherchons un système normal de coordonnées $(e_1 \ldots e_n)$ tel que, DANS CE SYSTÈME, *la matrice de* A *soit réduite de Schur*. La dernière colonne de la matrice réduite de l'opérateur A doit être

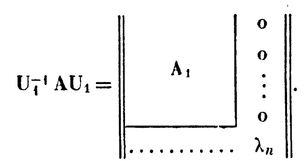
$$\left\| \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda_n \end{matrix} \right\|$$

Or, nous savons que les éléments de cette colonne sont les composantes (dans le système $e_1 \ldots e_n$) du vecteur transformé, par l'opérateur A, du dernier vecteur coordonné e_n; donc le système normal $(e_1 \ldots e_n)$ doit être tel que

$$A e_n = \lambda_n e_n ;$$

λ_n *est donc une valeur caractéristique de l'opérateur* A; et nous prendrons pour e_n une solution caractéristique correspondant à la racine λ_n. Nous avons ainsi déterminé le dernier vecteur coordonné du système coordonné par rapport auquel la matrice de A prendra la forme réduite de Schur; laissons, pour l'instant, arbitraires les vecteurs $e_1, e_2, \ldots, e_{n-1}$ en les astreignant toutefois à former avec e_n fixe un système normal provisoire de coordonnées; dans l'un quelconque de ces systèmes pro-

visoires la matrice de l'opérateur A aura la forme



Considérons l'opérateur qui, dans le système provisoire actuel, admet la matrice A_1, nous l'appellerons l'opérateur A_1 ; il opère dans le sous-espace défini par les vecteurs $e_1, e_2, \ldots, e_{n-1}$, qui est la projection de E parallèlement à e_n ; A_1 lie donc les projections, parallèlement à e_n, des deux vecteurs X et AX. Nous dirons que A_1 est *l'opérateur induit par* A *dans le sous-espace orthogonal à* e_n. *Dans ce sous-espace*, nous pouvons disposer arbitrairement du système orthogonal de vecteurs coordonnés $e_1, e_2, \ldots, e_{n-1}$ sans rien changer à la dernière colonne de la matrice de A ; nous allons profiter de cette latitude pour simplifier la matrice A_1, et, à cet effet, nous allons opérer comme précédemment : annuler tous les éléments de la dernière colonne, sauf le dernier. Nous serons conduit à la relation

$$A_1 e_{n-1} = \lambda_{n-1} e_{n-1},$$

d'où nous déduirons que λ_{n-1} doit être une valeur caractéristique du spectre de A_1. Mais nous avons

$$\text{dét}(A - \lambda) = (\lambda_n - \lambda)\,\text{dét}(A_1 - \lambda);$$

d'où il résulte immédiatement que les valeurs caractéristiques de A_1 sont les valeurs caractéristiques de A excepté λ_n. Nous prenons donc une valeur propre du spectre de A autre que λ_n, soit λ_{n-1} ; nous pouvons lui associer un vecteur e_{n-1} tel que, si $e_1, e_2, \ldots, e_{n-2}$ sont des vecteurs quelconques formant avec e_{n-1} et e_n déjà fixé un système normal provisoire de coordonnées, la matrice de l'opérateur A relative à ce second système provisoire ait la forme

$$U_2^{-1} A U_2 = \left\| \begin{array}{c|cc} A_2 & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \cdots\cdots\cdots & \lambda_{n-1} & 0 \\ \cdots\cdots\cdots & \cdots & \lambda_n \end{array} \right\|.$$

Et si nous continuons ainsi de proche en proche, nous arriverons finale-

ment à la forme réduite de Schur avec un système de vecteurs coordonnés e_n, e_{n-1}, ..., e_1 déterminé tel que, dans ce système, la matrice de A soit

$$U^{-1}AU = \begin{Vmatrix} \lambda_1 & 0 & 0 & \dots & 0 & 0 \\ \dots & \lambda_2 & 0 & \dots & 0 & 0 \\ \dots & \dots & \lambda_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \lambda_{n-1} & 0 \\ \dots & \dots & \dots & \dots & \dots & \lambda_n \end{Vmatrix}$$

tous les termes au-dessus de la diagonale principale étant nuls.

Remarquons que le vecteur e_n est invariant en direction par A; mais il n'en est pas de même de e_{n-1}. Le vecteur e_{n-1} est invariant par A_1; autrement dit *la projection de* Ae_{n-1} *sur l'espace orthogonal à* e_n *coïncide avec* e_{n-1}, ou bien encore, e_{n-1} est la *projection, sur l'espace orthogonal à* e_n, *d'un vecteur invariant en direction par* A *et correspondant à la valeur caractéristique* λ_{n-1}. Il est clair que nous aurions des résultats analogues pour les autres vecteurs coordonnés (¹).

2. La somme de deux matrices canoniques est une matrice canonique; en particulier, si A est une matrice canonique, il en est de même de $\lambda - A$. Il est évident, indépendamment de la démonstration que nous venons de faire, que les valeurs caractéristiques de la matrice réduite apparaissent dans sa diagonale principale. Si λ n'est pas une valeur caractéristique de la matrice canonique A, la matrice canonique $\lambda - A$ présente une inverse $(\lambda - A)^{-1}$ qui est *canonique* et non dégénérée, comme il est aisé de s'en assurer en résolvant le *système récurrent d'équations linéaires* symbolisé par

$$Y = (\lambda - A)X.$$

Considérons maintenant une matrice A non canonique; la substitution unitaire U qui la réduit à la forme canonique *réduit aussi sa résolvante* à sa forme canonique, car

$$U^{-1}\mathcal{R}_\lambda U = U^{-1}(\lambda - A)^{-1}U$$

s'écrit

$$U^{-1}\mathcal{R}_\lambda U = [U^{-1}(\lambda - A)U]^{-1};$$

par suite si $U^{-1}AU$ est canonique, $U^{-1}(\lambda - A)U$ l'est aussi; et si $U^{-1}(\lambda - A)U$ est canonique, $U^{-1}\mathcal{R}_\lambda U$ est aussi une matrice canonique,

(¹) *Voir* une solution algébrique de ce problème dans : SCHUR, *Math. Annalen*, 66, 1909; WINTNER, § 11, Chap. I.

puisque inverse de matrice canonique. Puisqu'il existe une même transformation unitaire réduisant simultanément à leurs formes canoniques respectives les deux matrices A et $\mathcal{R}_\lambda$, ces deux matrices sont dites *cogrédientes*. On peut encore dire que la résolvante de la réduite $U^{-1}AU$ d'une matrice A est la réduite $U^{-1}\mathcal{R}_\lambda U$ de la résolvante $\mathcal{R}_\lambda$ de A; car

$$(\lambda - U^{-1}AU)^{-1} = [U^{-1}(\lambda - A)U]^{-1}.$$

La forme canonique de Schur nous montre, de nouveau, que les valeurs caractéristiques d'une matrice sont invariantes dans une transformation unitaire puisqu'*elles apparaissent dans la diagonale de la forme canonique;* en particulier le déterminant de la matrice, qui est égal au produit des valeurs caractéristiques, est invariant dans une transformation unitaire.

Supposons que nous opérions sur une matrice A qui soit *hermitienne;* cette propriété se conservant par une transformation unitaire, la *forme canonique de Schur* devra être hermitienne, donc se réduira à une *matrice diagonale réelle*. Ce résultat est le même que celui obtenu par la méthode classique de réduction canonique à la forme diagonale d'une matrice hermitienne, ou, si l'on veut, à la ***réduction aux axes de la forme d'Hermite associée***. La méthode de réduction de Schur apparaît comme la généralisation de la méthode classique de réduction pour les matrices hermitiennes; dans le cas plus particulier des matrices réelles symétriques (et des formes quadratiques), la réduction est connue depuis longtemps, puisqu'elle est due à Lagrange et Gauss.

3. **Les opérateurs ou matrices normaux.** — Nous appellerons *opérateur* ou *matrice normal* un opérateur ou une matrice A permutable avec son associé, c'est-à-dire tel que

$$AA^* = A^*A.$$

Rappelons que $A^* = \overline{A'}$.

Les matrices ou opérateurs normaux forment une classe importante, étudiée par Schur et Töplitz, qui comprend comme cas particulier les opérateurs ou matrices d'Hermite définies par $A = A^*$, et les opérateurs ou matrices diagonaux (1).

Une transformation unitaire U *transforme une matrice normale en une autre matrice normale;* la vérification de cette proposition est

(1) Car l'associée d'une matrice diagonale est aussi diagonale et deux matrices diagonales sont toujours permutables.

immédiate : par hypothèse,

$$B = U^{-1}AU;$$

donc

$$BB^* = U^{-1}AUU^*A^*U^{-1*};$$

or, U étant unitaire, nous avons

$$UU^* = 1.$$

La relation précédente peut donc s'écrire

$$BB^* = U^{-1}AA^*U.$$

Et de même

$$B^*B = U^{-1}A^*AU.$$

Or $AA^* = A^*A$; il en résulte que $BB^* = B^*B$. Ceci prouve qu'il suffira d'avoir vérifié *dans un système particulier de coordonnées* que les matrices des opérateurs A et A^* sont permutables pour être sûrs *qu'elles le seront dans tout système* de coordonnées, c'est-à-dire que les opérateurs A et A^* seront permutables, donc A et A^* seront normaux. Nous avons signalé que les matrices diagonales sont des cas particuliers de matrices normales; montrons que toute matrice normale a pour réduite de Schur une matrice diagonale, ce qui revient à dire que *dans la géométrie unitaire toute matrice normale est équivalente à une matrice diagonale*. Une matrice normale, ayant pour réduite de Schur une matrice normale, il nous suffit d'établir qu'*une matrice, qui a la forme canonique de Schur et qui est normale*, *est nécessairement diagonale*.

Soit en effet $A = \| a_{ik} \|$ la matrice considérée; nous avons par hypothèse (forme canonique)

$$a_{ik} = 0 \qquad \text{pour} \qquad k > i.$$

Son associée est

$$A^* = \| \bar{a}_{ki} \|;$$

nous avons donc

$$AA^* = \left\| \sum_{j=1}^{n} a_{ij}\bar{a}_{kj} \right\|$$

et

$$A^*A = \left\| \sum_{j=1}^{n} a_{jk}\bar{a}_{ji} \right\|.$$

Exprimons que la matrice A est normale; nous obtenons

$$\sum_{j=1}^{n} a_{ij}\bar{a}_{kj} = \sum_{j=1}^{n} a_{jk}\bar{a}_{ji} \qquad \text{pour} \qquad i,\ k = 1,\ 2,\ \ldots,\ n.$$

Dans le cas particulier où $i = k$, cette formule se réduit à

$$\sum_{j=1}^{n} a_{ij}\bar{a}_{ij} = \sum_{j=1}^{n} a_{ji}\bar{a}_{ji},$$

soit

$$\sum_{j=1}^{i} | a_{ij} |^2 = \sum_{j=i}^{n} | a_{ji} |^2,$$

en tenant compte des termes nuls de la matrice A.

La somme des carrés des modules des éléments *de la ligne i ou de la colonne i* situés au-dessous de la diagonale principale est *la même*. Faisons d'abord i égal à 1; il vient

$$| a_{11} |^2 = \sum_{j=1}^{n} | a_{j1} |^2,$$

c'est-à-dire

$$\sum_{j=2}^{n} | a_{j1} |^2 = 0;$$

ce qui entraîne

$$a_{21} = a_{31} = \ldots = a_{n1} = 0.$$

Tous les éléments de la première colonne, situés sous la diagonale principale sont nuls. Si donc, au lieu de donner à i la valeur 1, nous lui donnons une valeur supérieure à 1, nous avons

$$\sum_{j=2}^{i} | a_{ij} |^2 = \sum_{j=i}^{n} | a_{ji} |^2.$$

En particulier, pour $i = 2$, nous obtenons

$$| a_{22} |^2 = \sum_{j=2}^{n} | a_{j2} |^2,$$

c'est-à-dire

$$\sum_{j=3}^{n} | a_{j2} |^2 = 0;$$

ce qui entraîne

$$a_{32} = a_{42} = \ldots = a_{n2} = 0.$$

Tous les éléments de la deuxième colonne, situés sous la diagonale principale sont nuls. Si donc, au lieu de donner à i la valeur 2, nous lui donnons une valeur supérieure à 2, nous avons

$$\sum_{l=3}^{i} |a_{il}|^2 = \sum_{l=i}^{n} |a_{li}|^2.$$

En particulier, pour $i = 3$, ... Il est clair, qu'en opérant ainsi de proche en proche, nous démontrerons que tous les termes situés sous la diagonale principale sont nuls; or, par hypothèse, ceux situés au-dessus de la diagonale principale le sont également; il en résulte que la matrice A présente bien la forme diagonale. Nous pouvons dire, en abrégé, que les matrices *diagonales* sont des *normales réduites*.

La proposition précédente nous permet d'affirmer qu'*il existe bien des matrices qui ne sont pas normales;* il suffit d'envisager une matrice de Schur ayant des termes non nuls au-dessous de sa diagonale principale; par exemple,

$$A = \begin{Vmatrix} 0 & 0 \\ 1 & 0 \end{Vmatrix}.$$

Vérifions, comme exercice, que cette matrice n'est pas normale; à cet effet cherchons son associée; c'est

$$A^* = \begin{Vmatrix} 0 & 1 \\ 0 & 0 \end{Vmatrix}.$$

Et nous constatons bien que les deux matrices AA^* et A^*A ne sont pas égales puisque

$$AA^* = \begin{Vmatrix} 0 & 0 \\ 0 & 1 \end{Vmatrix},$$

$$A^*A = \begin{Vmatrix} 1 & 0 \\ 0 & 0 \end{Vmatrix}.$$

4. Indiquons quelques propriétés évidentes des *matrices équivalentes en géométrie unitaire*, dont nous avons déjà rencontré les analogues à l'occasion des *matrices équivalentes en géométrie homogène linéaire* (*voir* Chap. II, § II).

Deux matrices A et B sont équivalentes dans la géométrie unitaire s'il est possible de trouver une matrice unitaire U telle que

$$B = U^{-1}AU.$$

Si deux matrices sont équivalentes, elles le demeurent après une trans-

formation unitaire V; en effet,

$$V^{-1}BV = V^{-1}U^{-1}AUV = V^{-1}U^{-1}VV^{-1}AVV^{-1}UV;$$

soit

$$(V^{-1}BV) = [V^{-1}UV]^{-1}(V^{-1}AV)[V^{-1}UV];$$

or la matrice $V^{-1}UV$, produit de trois matrices unitaires, est unitaire; ce qui démontre la proposition.

L'ensemble des matrices équivalentes à une matrice A, forme une *classe de matrices équivalentes dans la géométrie unitaire*, car deux matrices quelconques de l'ensemble sont équivalentes; nous avons par hypothèse

$$B = U^{-1}AU, \qquad \text{soit} \qquad A = UBU^{-1},$$
$$C = V^{-1}AV;$$

d'où il résulte que

$$C = V^{-1}UBU^{-1}V = [U^{-1}V]^{-1}B[U^{-1}V];$$

et il suffit de remarquer alors que la matrice $U^{-1}V$ est unitaire. Il est aussi évident qu'une matrice équivalente à une matrice de la classe appartient à la classe.

Nous pouvons encore dire qu'une classe de matrices équivalentes est constituée par *l'ensemble des matrices représentant un même opérateur linéaire dans tous les systèmes normaux de coordonnées.* Si deux matrices A et B sont équivalentes, elles auront *une* même matrice réduite de Schur et réciproquement. En particulier, si deux matrices *normales* A et B sont équivalentes, elles auront *une* même matrice *diagonale* réduite de Schur et réciproquement; mais les éléments de la diagonale de la matrice de Schur sont les valeurs caractéristiques, donc deux matrices normales équivalentes ont même valeurs caractéristiques. Nous allons montrer que, réciproquement, *deux matrices normales* A *et* B *ayant mêmes valeurs caractéristiques sont équivalentes dans le groupe unitaire.* En effet, leurs matrices réduites de Schur sont des *matrices diagonales ayant mêmes éléments à l'ordre près;* il nous suffit donc de prouver que par une (ou plusieurs) transformation unitaire, nous pouvons permuter, de telle façon que nous le désirons, les éléments d'une matrice diagonale.

En opérant de proche en proche, comme on le fait par exemple en arithmétique pour démontrer que le produit de plusieurs nombres est commutatif, nous sommes ramenés à démontrer que par une transformation unitaire il est possible de permuter *deux éléments consécutifs de la diagonale principale* d'une matrice diagonale; prenons les deux

premiers éléments, pour fixer les idées, et montrons donc que les deux matrices

$$A = \begin{Vmatrix} \lambda_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_4 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{Vmatrix} \quad \text{et} \quad B = \begin{Vmatrix} \lambda_2 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_4 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{Vmatrix}$$

sont équivalentes dans le groupe unitaire. Démontrons qu'il existe *une matrice de permutation* P transformant la matrice A en la matrice B. Nous désignons, sous le nom de « matrice de permutation », une matrice qui a *dans chaque ligne et dans chaque colonne un élément non nul et un seul*, cet élément étant égal à 1; les matrices de permutation vérifient

$$PP' = 1,$$

car l'élément général de la matrice produit est

$$\sum_l a_{il} a_{kl};$$

et puisque dans une colonne il n'y a qu'un élément non nul, a_{il} et a_{kl} ne pourront être simultanément non nuls que si $i = k$; dans ce cas, cela ne sera réalisé que pour un nombre l, leur donnant à tous deux la valeur 1; soit

$$\Sigma a_{il} a_{kl} = \delta_{ik}.$$

De la relation

$$PP' = 1$$

nous déduisons

$$PP^* = 1,$$

puisque la matrice P' est réelle, c'est dire $P' = \overline{P'}$; *les opérateurs de permutation sont donc des cas particuliers des opérateurs unitaires.* Envisageons alors

$$P = \begin{Vmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 \end{Vmatrix}$$

nous voyons que

$$P = P',$$

ce qui entraîne

$$P = P^{-1};$$

nous obtenons alors successivement

$$AP^{-1} = \begin{Vmatrix} 0 & \lambda_1 & 0 & 0 & \dots & 0 \\ \lambda_2 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_4 & \dots & 0 \\ . & . & . & .. & \dots & . \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{Vmatrix}.$$

$$PAP^{-1} = \begin{Vmatrix} \lambda_2 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_4 & \dots & 0 \\ . & . & . & .. & \dots & . \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{Vmatrix} = B.$$

C. Q. F. D.

Donc *si* A *et* B *normales ont même spectre*, leurs réduites $U^{-1}AU$ et $V^{-1}BV$ sont équivalentes par une matrice de permutation (donc unitaire)

$$U^{-1}AU = PV^{-1}BVP^{-1},$$

soit

$$A = UPV^{-1}BVP^{-1}U^{-1} = (VP^{-1}U^{-1})^{-1}B(VP^{-1}U^{-1});$$

d'où il résulte que *les matrices* A *et* B *sont équivalentes dans la géométrie unitaire.*

Conclusion. — *Le spectre d'une matrice normale est un système complet d'invariants par rapport au groupe unitaire, qui caractérise la classe de cette matrice dans le groupe unitaire.*

5. Résolvantes des matrices normales. Spectre. Solutions propres. — Nous avons montré qu'une même substitution linéaire réduit la matrice A et sa résolvante $\mathcal{R}_\lambda = (\lambda - A)^{-1}$ à leurs formes canoniques; nous allons donc prendre la matrice A sous sa forme normale qui est diagonale, soit

$$A = \| \delta_{ik}\lambda_i \|,$$

les nombres λ_i étant les valeurs propres; nous en déduisons immédiatement

$$\lambda - A = \| (\lambda - \lambda_i)\delta_{ik} \|$$

et

$$\mathcal{R}_\lambda = (\lambda - A)^{-1} = \left\| \frac{\delta_{ik}}{\lambda - \lambda_i} \right\|.$$

Réciproquement, si $\mathcal{R}_\lambda$ présente cette forme, la matrice A est diagonale donc normale réduite.

La résolvante $\mathcal{R}_\lambda$ a les *pôles simples* $\lambda = \lambda_i$; il est clair que cette proposition se conserve par toute transformation unitaire déterminée par une matrice U, autrement dit $U^{-1}\mathcal{R}_\lambda U$ a aussi tous ses pôles simples. Par conséquent, si la matrice A a une résolvante à pôles simples, il en est de même de la matrice $U^{-1}AU$. Finalement, nous voyons que *toutes les matrices normales ont des résolvantes à pôles simples.*

Plusieurs valeurs du spectre peuvent être confondues; désignons par $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p)}$ les valeurs propres distinctes, ayant respectivement pour ordre de multiplicité $s_1, s_2, \ldots, s_p$; et considérons l'équation

$$(1) \qquad (\lambda^{(j)} - A)X = 0,$$

relative à une matrice A *normale*. Nous savons que cette équation présente des solutions X non nulles; nous allons voir que *dans le cas des matrices normales elle présente s_j solutions linéairement indépendantes*, s_j désignant l'ordre de multiplicité de la racine caractéristique $\lambda^{(j)}$. Nous pouvons toujours supposer la matrice A sous forme canonique, ce qui revient simplement à choisir convenablement le système de coordonnées avec lequel nous allons opérer. L'équation précédente devient alors

$$\|(\lambda^{(j)} - \lambda_i)\delta_{ik}\| X = 0,$$

ce qui résume symboliquement le système d'équations

$$(2) \qquad (\lambda^{(j)} - \lambda_i)x_i = 0 \quad \text{pour} \quad i = 1, 2, \ldots, n.$$

Pour les valeurs de i correspondant à $\lambda_i \neq \lambda^{(j)}$ nous aurons $x_i = 0$; pour les valeurs de i correspondant à $\lambda_i = \lambda^{(j)}$, nous aurons x_i arbitraire; or ces dernières valeurs x_i sont au nombre de s_j. Les solutions du système d'équations (2) dépendent donc bien de s_j constantes arbitraires; ce système présente *s_j solutions élémentaires linéairement indépendantes*, $X_j^{(1)}, X_j^{(2)}, \ldots, X_j^{(s_j)}$, dont les coordonnées se trouvent dans les colonnes du tableau suivant :

$$X_j^{(1)}, \quad X_j^{(2)}, \quad X_j^{(3)}, \quad \ldots, \quad X_j^{(s_j)},$$

coordonnées correspondant :

Aux indices i pour lesquels $\lambda_i \neq \lambda^{(j)}$:

$$\begin{matrix} 0 & 0 & 0 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 0 \end{matrix}$$

indices i pour lesquels $\lambda_i = \lambda^{(j)}$:

$$\begin{matrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 1 \end{matrix}$$

indices i pour lesquels $\lambda_i \neq \lambda^{(j)}$:

$$\begin{matrix} 0 & 0 & 0 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 0 \end{matrix}$$

Il est clair que ces solutions sont linéairement indépendantes, et que l'ensemble de toutes les solutions obtenues de la même manière en donnant successivement à j les valeurs $1, 2, \dots, p$ sont linéairement indépendantes.

Nous voyons donc que, *pour une matrice normale*, à un nombre quelconque de valeurs propres *distinctes ou non* correspond le même nombre de solutions propres linéairement indépendantes.

Toutes les propriétés précédentes relatives aux matrices normales, s'appliquent en particulier aux matrices hermitiennes, dont nous ferons une étude détaillée dans le prochain chapitre, et elles expliquent les propriétés particulières de ces matrices.

6. Normes matricielles d'une matrice quelconque. — Introduites par Schur, les normes matricielles différencient encore les matrices normales des matrices non normales.

Nous avons désigné par norme d'un vecteur X (matrice à une seule colonne) l'expression

$$X^*X = |X|^2.$$

Nous généralisons et appelons *normes matricielles d'une matrice* A, les deux matrices A^*A et AA^*; la première est dite *norme antérieure* et la seconde *norme postérieure*. Les deux normes d'une matrice quel-

conque sont des matrices *hermitiennes*, donc normales, puisque

$$(AA^*)^* = AA^* \quad \text{et} \quad (A^*A)^* = A^*A.$$

Si la matrice A est normale, ses deux normes sont égales; si la matrice A n'est pas normale, ses deux normes ne sont pas égales. Rappelons les deux formules suivantes démontrées antérieurement (*voir* Chap. 3, § I, n° 9).

$$AA^*(X, X) = X^*AA^*X = \sum_{k=1}^{n} \left| \sum_{i=1}^{n} a_{ik} x_i \right|^2$$

$$A^*A(X, X) = X^*A^*AX = \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} \bar{x}_k \right|^2$$

Les deux formes d'Hermite associées aux normes matricielles sont donc définies positives.

Nous avons aussi démontré

$$\text{Trace}(AA^*) = \text{Trace}(A^*A) = \sum_{ik} |a_{ik}|^2,$$

cette valeur commune s'appelle la *trace normale* (normalspur) de la matrice A.

La trace normale d'une matrice A étant invariante par une transformation unitaire ([1]), nous pouvons la calculer *sur la forme réduite de Schur* $\|\alpha_{ik}\|$; il devient alors évident que :

$$\text{Trace}(AA^*) = \sum_{i,k} |a_{ik}|^2 > \sum_{i} |\alpha_{ii}|^2 = \sum_{i} |\lambda_i|^2 \qquad \text{si A n'est pas normale;}$$

$$\text{Trace}(AA^*) = \sum_{i,k} |a_{ik}|^2 = \sum_{i} |\alpha_{ii}|^2 = \sum_{i} |\lambda_i|^2 \qquad \text{si A est normale.}$$

Nous allons montrer que les deux normes AA^* et A^*A d'une matrice quelconque A sont toujours *équivalentes dans le groupe unitaire;* puisque les normes sont des matrices hermitiennes, donc normales, il nous suffit, pour cela, de prouver qu'elles ont même spectre. Soit à établir que les deux équations

$$\text{dét}(\lambda - AA^*) = 0 \quad \text{et} \quad \text{dét}(\lambda - A^*A) = 0,$$

ont mêmes racines. Envisageons d'abord le cas général où la matrice A

(1) *Voir* Chap. II, § IV, n° 5.

n'est pas dégénérée; elle présente alors une inverse et nous avons

$$A^{-1}AA^*A = A^*A,$$

ce qui nous prouve que la norme A^*A est transformée de la norme AA^* par la matrice A; d'après une propriété classique, elles ont alors même polynome caractéristique (*voir* Chap. II, § II, n° 4)

$$\text{dét}(\lambda - AA^*) = \text{dét}(\lambda - A^*A).$$

Si la matrice A est dégénérée, nous pouvons nous ramener aisément au cas précédent. En effet $\lambda = 0$ est une valeur propre *isolée* de la matrice A, puisque les valeurs propres sont en nombre fini. Si nous choisissons ε quelconque, mais de module suffisamment petit et non nul, la matrice

$$B = \varepsilon - A$$

ne sera pas dégénérée; et nous aurons par conséquent

$$\text{dét}(\lambda - BB^*) = \text{dét}(\lambda - B^*B);$$

cette relation étant vérifiée quelque petit que soit ε, nous obtenons, en passant à la limite

$$\text{dét}(\lambda - AA^*) = \text{dét}(\lambda - A^*A).$$

Nous ne nous étendrons pas plus sur l'étude des matrices normales, on pourra compléter les résultats donnés en consultant le Mémoire de Schur (*Math. Annalen*, t. 66, 1909).

CHAPITRE IV.

ÉTUDE PARTICULIÈRE DES MATRICES ET OPÉRATEURS HERMITIENS.

I. — RÉDUCTION.

1. Les matrices hermitiennes sont les matrices égales à leur associée, soit $A = A^*$; nous avons donc

$$AA^* = A^*A = A^2.$$

La matrice A^2 est aussi hermitienne puisque $(AA^*)^* = AA^*$.

Spectre. — En remarquant que toute matrice équivalente dans le groupe unitaire à une matrice hermitienne est aussi une matrice hermitienne, nous avons démontré que la matrice réduite de Schur d'une matrice *hermitienne est diagonale et a tous ses termes réels et réciproquement.* Nous pouvons dire que les matrices hermitiennes sont les *matrices normales à spectre réel.*

Espaces propres. — Soient donc $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p)}$ les racines distinctes de l'équation caractéristique de l'opérateur hermitien A; elles sont réelles et ont respectivement comme ordre de multiplicité s_1, $s_2, \ldots s_p$. L'équation

$$(\lambda^{(j)} - A)X = 0$$

définit un espace vectoriel $E^{(j)}$ à s_j dimensions, ainsi que nous l'avons montré dans le Chapitre précédent. Il est clair que ces espaces $E^{(1)}$, $E^{(2)}, \ldots, E^{(p)}$ sont *invariants* par l'opérateur A. Montrons que ces espaces sont deux à deux *orthogonaux*; soient en effet un vecteur X' de l'espace $E^{(j)}$ et un vecteur X'' de l'espace $E^{(k)}$; nous avons

$$(\lambda^{(j)} - A)X' = 0,$$
$$(\lambda^{(k)} - A)X'' = 0,$$

ce qui peut s'écrire

$$AX' = \lambda^{(j)} X',$$
$$AX'' = \lambda^{(k)} X''.$$

Multiplions scalairement à droite la première de ces relations par X'' et à gauche la seconde par X'; retranchons. Il vient

$$(AX', X'') - (X', AX'') = \left[\overline{\lambda^{(j)}} - \lambda^{(k)}\right](X', X'').$$

L'opérateur A étant hermitien le premier membre est nul (*voir* n° 8, § I, Chap. III). Il en résulte; puisque $\overline{\lambda^{(j)}} = \lambda^{(j)}$ réel est différent de $\lambda^{(k)}$, que

$$(X', X'') = 0.$$

Les sous-espaces propres sont donc deux à deux orthogonaux; en particulier ils sont *disjoints*. Puisque la somme de leurs nombres de dimensions $s_1 + s_2 + \ldots + s_p$ est égale à n, ils forment donc une décomposition de l'espace

$$E = E^{(1)} + E^{(2)} + \ldots + E^{(p)}.$$

Les sous-espaces étant orthogonaux nous pouvons adapter à cette décomposition de l'espace un système de coordonnées normal formé par la réunion de systèmes de coordonnées normaux quelconques de chaque sous-espace.

Les sous-espaces étant invariants, la matrice réduite de A se réduit à une suite de p matrices situées le long de sa diagonale principale, ayant respectivement pour ordre $s_1, s_2, \ldots, s_p$ et représentant l'opérateur induit par A dans le sous-espace correspondant. Or, dans le sous-espace $E^{(i)}$, nous avons

$$AX = \lambda^{(i)} X.$$

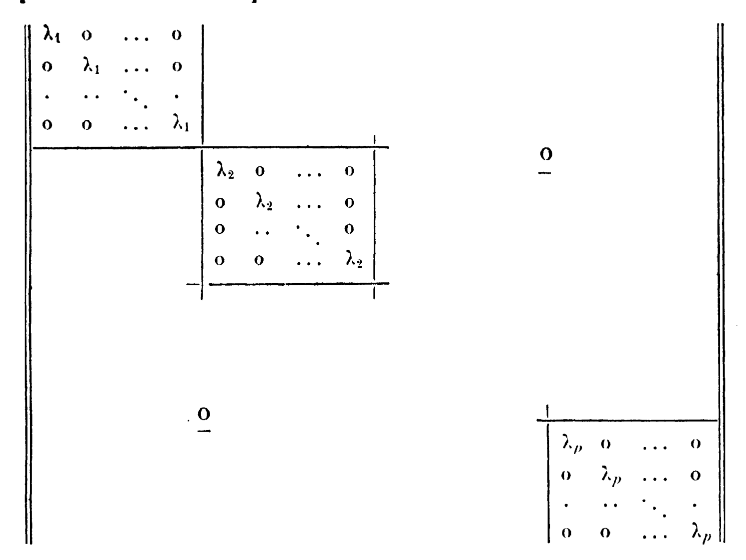
Il en résulte que l'opérateur induit A, se réduit à un *opérateur scalaire de multiplication*

$$\left\| \begin{matrix} \lambda^{(i)} & 0 & 0 & \ldots & 0 \\ 0 & \lambda^{(i)} & 0 & \ldots & 0 \\ 0 & 0 & \lambda^{(i)} & \ldots & 0 \\ . & . & \ldots & \ldots & . \\ 0 & 0 & 0 & \ldots & \lambda^{(i)} \end{matrix} \right\|,$$

quel que soit le système de coordonnées normal choisi dans $E^{(i)}$.

Finalement, relativement à un système de coordonnées normal quelconque adapté à la décomposition de l'espace E en les sous-espaces $E^{(i)}$,

l'opérateur hermitien A possède la matrice réduite



Nous avons donc un certain arbitraire dans le choix du système d'axes donnant la forme réduite. Toutefois, si les racines caractéristiques étaient simples, les axes de coordonnées seraient imposés (à une permutation près).

Par définition, l'équation

$$AX = \alpha X,$$

définit le sous-espace $E^{(\alpha)}$ *propre à* α; si α n'est pas une valeur caractéristique le sous-espace propre correspondant est vide; si α est une valeur caractéristique λ_i le sous-espace propre correspondant est non vide et n'est autre que le sous-espace $E^{(i)}$.

2. Envisageons la *forme d'Hermite associée* à la matrice A; elle s'écrit

$$A(X, X) = X^* AX = \sum_{i,k=1}^{n} a_{ik} \overline{x_i} x_k.$$

Si nous effectuons le changement d'axes unitaire défini par la matrice unitaire U

$$X = U\Xi,$$

la forme d'Hermite devient

$$X^* AX = \Xi^* U^* AU\Xi.$$

Elle correspond donc dans les nouveaux axes à la matrice

$$U^* AU = U^{-1} AU.$$

Choisissons pour U la transformation unitaire ramenant la matrice A à sa forme canonique $\|\lambda_i \delta_{ik}\|$; la forme d'Hermite associée sera également ramenée à sa forme canonique

$$\sum_{i=1}^{n} \lambda_i \bar{\xi}_i \xi_i.$$

Étant donné un opérateur hermitien A, nous pouvons donc toujours le rapporter à des axes de coordonnées normales e_i (*axes principaux*) tels que, pour $X = x_1 e_1 + \ldots + x_n e_n$, on ait simultanément

$$A(X, X) = \sum_{i=1}^{n} \lambda_i \bar{x}_i x_i,$$

$$(X; X) = \sum_{i=1}^{n} \bar{x}_i x_i;$$

les λ_i sont les *valeurs propres* de l'opérateur A; les vecteurs coordonnés sont les *vecteurs propres* de l'opérateur.

Désignons par X_i la projection orthogonale du vecteur X sur le sous-espace $E^{(i)}$ et posons

$$E_{\lambda^{(i)}}(X) = (X_i, X_i).$$

On a par exemple

$$X_1 = x_1 e_1 + \ldots + x_{s_1} e_{s_1},$$

$$E_{\lambda^{(1)}}(X) = \bar{x}_1 x_1 + \ldots + \bar{x}_{s_1} x_{s_1} = (X_1, X_1).$$

Il résulte de cette définition que nous avons

$$(X, X) = \sum_{\alpha} E_\alpha(X),$$

$$A(X, X) = \sum_{\alpha} \alpha E_\alpha(X),$$

formules dans lesquelles α prend successivement les valeurs $\lambda^{(1)}$, $\lambda^{(2)}$, ..., $\lambda^{(p)}$, ou bien encore où α prend toutes les valeurs en faisant la convention

$$E_\alpha(X) = 0 \qquad \text{si } \alpha \text{ est différent de tous les } \lambda^{(i)},$$

$$E_\alpha(X) = E_{\lambda^{(i)}}(X) \qquad \text{si } \alpha \text{ est égal à } \lambda^{(i)}.$$

Considérons les projections orthogonales $X_1, X_2, \ldots, X_p$ d'un vecteur X sur les sous-espaces $E^{(1)}, E^{(2)}, \ldots, E^{(p)}$; X_i sera obtenu à partir de X par un *opérateur* E_i *de projection* tel que

$$X_i = E_i X.$$

Ces opérateurs E_i sont en même nombre que les valeurs caractéristiques *distinctes;* dans le système canonique de coordonnées l'opérateur E_i possède la matrice E_i

$$E_i = \left\| \begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & \begin{matrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ & & & \ddots \\ 0 & & & & 1 \end{matrix} & 0 \\ \hline 0 & 0 & 0 \end{array} \right\|$$

dite *matrice de projection* (*Einzelmatrix* des Allemands). Elle se déduit de la matrice unité en ne conservant que les termes des lignes et colonnes conduisant dans la matrice réduite de A, à la valeur caractéristique $\lambda^{(i)}$ dans la diagonale principale, les autres termes étant remplacés par zéro. Nous avons évidemment les deux relations matricielles suivantes :

$$1 = \sum_{\alpha} E_{\alpha},$$

$$A = \sum_{\alpha} \alpha E_{\alpha}.$$

Les E_i n'admettent évidemment que les valeurs propres 0 et 1. Et il est clair que $E_i^2 = E_i$. Nous verrons que la réciproque est vraie.

3. Tout opérateur de projection sur un sous-espace quelconque $E^{(1)}$ est un opérateur *hermitien*. Décomposons, en effet, l'espace E en deux sous-espaces complémentaires $E^{(1)}$ et $E^{(2)}$ *orthogonaux*, un vecteur de E sera la somme de deux vecteurs appartenant chacun à l'un des sous-espaces

$$X = X_1 + X_2 \quad \text{avec} \quad X_1 = E_1 X,$$
$$Y = Y_1 + Y_2 \quad \text{avec} \quad Y_1 = E_1 Y.$$

Nous avons

$$(X_1, Y) = (X_1, Y_1 + Y_2) = (X_1, Y_1) + (X_1, Y_2) = (X_1, Y_1)$$

et de même

$$(X, Y_1) = (X_1 + X_2, Y_1) = (X_1, Y_1) + (X_2, Y_1) = (X_1, Y_1),$$

d'où

$$(X_1, Y) = (X, Y_1).$$

Soit

$$(E_1 X, Y) = (X, E_1 Y),$$

relation qui, vérifiée pour X et Y quelconque, caractérise un opérateur hermitien. D'autre part $X^* E_1 X$ est le carré de la longueur du vecteur X_1,

Désignons par E_2 l'opérateur de projection faisant passer de E à $E^{(2)}$, autrement dit $1 - E_1$, ($X_2 = E_2 X$ donne $X = E_1 X + E_2 X$, c'est-à-dire $E_2 = 1 - E_1$). Les sous-espaces $E^{(1)}$ et $E^{(2)}$ étant orthogonaux nous avons

$$E_1 E_2 = 0;$$

et de même $E_2 E_1 = 0$ qui pourrait se déduire de la première relation en prenant son associée et en remarquant qu'un opérateur hermitien est égal à son associé. Les opérateurs E_1 et E_2 sont dits *indépendants*. Cette relation peut encore s'écrire

$$E_1(1 - E_1) = 0$$

ou bien

$$E_1 = E_1^2,$$

de même

$$E_2 = E_2^2.$$

Si un espace est décomposé en *p* espaces deux à deux orthogonaux $E^{(1)}$, $E^{(2)}$, ..., $E^{(p)}$, les opérateurs de projection correspondants étant E_1, E_2, ..., E_p, nous aurons

$$X = E_1 X + E_2 X + \ldots + E_p X,$$

et l'on vérifie aisément

$$E_i E_j = \delta_{ij} E_i.$$

Montrons que, *réciproquement, si un opérateur hermitien* E_1 *satisfait à* $E_1^2 = E_1$, *c'est un opérateur de projection.*

En effet l'opérateur $E_2 = 1 - E_1$ est aussi hermitien et satisfait à la même relation, car

$$E_2^2 = (1 - E_1)^2 = 1 - 2E_1 + E_1^2 = 1 - E_1 = E_2.$$

D'autre part, nous avons

$$E_1 E_2 = E_1(1 - E_1) = E_1 - E_1^2 = 0.$$

Les deux opérateurs E_1 et E_2 sont donc indépendants. De plus, puisque

$$1 = E_1 + E_2,$$

tout vecteur X est décomposé en une somme de deux vecteurs

$$X = E_1 X + E_2 X,$$

appartenant aux deux sous-espaces

$$E^{(1)} = E_1 X \quad \text{pour X arbitraire,}$$
$$E^{(2)} = E_2 Y \quad \text{pour Y arbitraire.}$$

Ces deux sous-espaces sont orthogonaux (donc disjoints), car

$$(E_1 X, E_2 Y) = (X, E_1 E_2 Y),$$

puisque E_1 est hermitien et cette expression est nulle en vertu de la relation $E_1 E_2 = 0$.

La décomposition $X = E_1 X + E_2 X$ comprend donc un vecteur $E_1 X$ dans un sous-espace E^1 et un vecteur $E_2 X$ orthogonal à ce sous-espace; $E_1 X$ est donc la projection orthogonale de X sur E^1; E_1 est bien un opérateur de projection.

Plus généralement, si les opérateurs hermitiens $E_1, E_2, \ldots, E_p$ sont indépendants, c'est-à-dire tels que $E_i E_i = E_i$ et $E_i E_j = 0$ si $i \neq j$, ils définissent des multiplicités $E^{(i)} = E_i X$ deux à deux orthogonales; si de plus ces multiplicités permettent d'engendrer tout l'espace, c'est-à-dire si $E = E^{(1)} + E^{(2)} + \ldots + E^{(n)}$, nous aurons

$$X = E_1 X + E_2 X + \ldots + E_n X,$$

et l'opérateur E_i définira la projection orthogonale de l'espace E sur le sous-espace $E^{(i)}$.

En résumé, *toute forme, matrice ou opérateur d'Hermite* A *fait correspondre aux nombres réels* α, *des opérateurs, formes ou matrices de projection* E_α, *indépendants, tels que*

$$1 = \sum_\alpha E_\alpha,$$

$$A = \sum_\alpha \alpha E_\alpha,$$

les E_α *n'étant différents de zéro que pour un nombre fini de valeurs de* α qui sont les valeurs caractéristiques λ_i de A.

Faisons maintenant une remarque importante. Considérons la relation

$$A = \sum_\alpha \alpha E_\alpha$$

et élevons-la au carré; du fait que les opérateurs E_α sont indépendants, il vient immédiatement

$$A^2 = \sum_\alpha \alpha^2 E_\alpha$$

Nous pouvons multiplier maintenant cette seconde relation par la première et ainsi de suite. De proche en proche, nous arrivons ainsi à la relation générale

$$A^h = \sum_\alpha \alpha^h E_\alpha.$$

Autrement dit les différentes puissances d'une matrice hermitienne donnent naissance *aux mêmes matrices de projection* et ont pour valeurs caractéristiques les puissances des valeurs caractéristiques de la première.

Généralisons de nouveau : envisageons un polynome en A à coefficients *réels*, soit $f(A)$

$$f(A) = c_0 1 + c_1 A + \ldots + c_h A^h.$$

Nous aurons

$$f(A) = c_0 \sum_\alpha E_\alpha + c_1 \sum_\alpha \alpha E_\alpha + \ldots + c_h \sum_\alpha \alpha^h E_\alpha,$$

soit

$$f(A) = \sum_\alpha f(\alpha) E_\alpha.$$

L'opérateur hermitien $f(A)$, (f *polynome*) *donne donc naissance aux mêmes matrices de projection que* A; *ses valeurs caractéristiques sont* $f(\lambda_i)$, *si nous désignons par* λ_i *les valeurs caractéristiques de* A.

4. Valeurs de la forme attachée à un opérateur hermitien A. — Considérons la forme $A(X, X)$ attachée à l'opérateur A

$$A(X, X) = X^* A X = \sum_{i,k} a_{ik} \overline{x_i} x_k.$$

Nous avons vu qu'il est possible de prendre des axes de coordonnées e (axes principaux) tels que la forme se réduise à

$$A(X, X) = \sum_{i=1}^{n} \lambda_i \bar{x}_i x_i,$$

où x_i sont les nouvelles coordonnées du vecteur X, définies par la relation vectorielle :

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n,$$

$e_1, e_2, \ldots, e_n$ formant un système de vecteurs orthonormaux.

Nous nous proposons de chercher le domaine décrit dans le plan complexe par la valeur de la fonction A(X, X) du vecteur X, lorsque celui-ci varie d'une façon quelconque.

Tous les λ_i étant réels, il est évident que A(X, X) sera toujours réel. Si nous supposons maintenant que toutes les valeurs spectrales λ_i *sont positives*, donc supérieures à un nombre positif λ, il en résulte que

$$A(X, X) = \sum_{i=1}^{n} \lambda_i \bar{x}_i x_i > \lambda \sum_{i=1}^{n} \bar{x}_i x_i = \lambda |X|^2;$$

autrement dit la forme A est toujours positive sauf pour $X = 0$, auquel cas elle est nulle. Nous dirons alors que la forme d'Hermite est *définie positive*. Si, en particulier, *nous astreignons le vecteur* X *à la condition* $|X| = 1$, nous aurons

$$A(X, X) > \lambda > 0.$$

Supposons que tous les λ_i soient *positifs ou nuls, l'un d'entre eux au moins étant nul*. Désignons par $\lambda^{(p)} = 0$ l'ensemble des valeurs propres nulles, il vient

$$A(X, X) = \sum_{i=1}^{n} \lambda_i \bar{x}_i x_i > \lambda \sum_{i=1}^{n}{}' \bar{x}_i x_i,$$

en désignant par λ un nombre inférieur à tous les λ_i non nuls et par Σ' une somme étendue à tous les indices i compris entre 1 et n et ne correspondant pas à une valeur caractéristique nulle. Il est clair alors que la forme A est positive ou nulle; elle est nulle dans le cas, et seulement dans le cas, où le vecteur X est choisi dans la multiplicité $E^{(p)}$. Remarquons principalement que cette forme prend des valeurs nulles pour des vecteurs non nuls; elle est dite *définie non négative*.

Nous pourrions définir de même des formes d'Hermite *définies négatives* et des formes d'Hermite *définies non positives*.

Étudions, plus généralement le domaine décrit par une forme d'Hermite $A(X, X)$ quelconque, *lorsque nous astreignons le vecteur* X *à la condition* $|X| = 1$. Nous avons

$$A(X, X) = \sum_{i=1}^{n} \lambda_i \bar{x}_i x_i = \sum_{i=1}^{n} \lambda_i |x_i|^2.$$

Soit λ la plus petite valeur caractéristique et Λ la plus grande. Les points représentatifs des nombres λ_i sont distribués sur le segment (λ, Λ) extrémités comprises; si nous les affectons des masses positives $|x_i|^2$ correspondantes, le centre de gravité de ce système a pour abscisse

$$A(X, X) = \sum_{i=1}^{n} \lambda_i |x_i|^2$$

puisque

$$\sum_{i=1}^{n} |x_i|^2 = 1.$$

Il en résulte que le point représentant la valeur de la forme $A(X, X)$ *décrit le segment* (λ, Λ) lorsque X varie arbitrairement sur la sphère $|X| = 1$:

$$\lambda \leqq A(X, X) \leqq \Lambda.$$

Le nombre λ, *minimum absolu de la forme*, est atteint pour $X = e_i$ si $\lambda = \lambda_i$ et le nombre Λ, *maximum absolu de la forme* pour $X = e_j$ si $\Lambda = \lambda_j$. Le domaine décrit par la forme est donc *le plus petit segment qui contient le spectre*. Dans la suite, nous généraliserons ce résultat.

5. **Applications.** — A étant une matrice hermitienne, les valeurs propres de A^2 sont égales aux carrés de celles de A. Il en résulte que la forme A^2 est toujours *définie non négative*. Nous pouvons même dire que : la condition nécessaire et suffisante pour que la matrice A ne soit pas dégénérée est que la forme A^2 soit *définie positive* (¹). La réciproque peut encore s'exprimer de la façon suivante :

Toute matrice d'Hermite A non négative est le carré d'une matrice d'Hermite B. En effet, ramenons A à la forme diagonale

$$A = \| \lambda_i \quad \delta_{ik} \|;$$

(¹) Cette propriété s'étend aux matrices infinies (Töplitz).

tous les λ_i sont positifs ou nuls. Posons alors

$$B = \| \sqrt{\lambda_i} \quad \delta_{ik} \|;$$

B est une forme d'Hermite et il est évident que $A = B^2$.

Si la forme A^2 est définie positive, la matrice A n'est pas dégénérée et par conséquent la matrice A^{-1} existe; ses valeurs propres étant λ_i^{-1} il en résulte que *la matrice* A^{-1} *est définie positive ou définie négative en même temps que la matrice* A.

6. Nous avons déjà montré que la condition nécessaire et suffisante pour que le produit AB de deux matrices hermitiennes A et B soit hermitien est que ces deux matrices soient *permutables*, c'est-à-dire que l'on ait

$$AB = BA.$$

La condition nécessaire et suffisante pour qu'il en soit ainsi est que A *et* B *soient réductibles à la forme diagonale par la même substitution unitaire* U, *autrement dit que les formes d'Hermite* A(X, X) *et* B(X, X) *aient même axes principaux.*

La condition est suffisante. En effet, il nous suffit de vérifier que les matrices sont permutables dans un système particulier de coordonnées; nous choisirons évidemment le système de coordonnées par rapport auquel les matrices prennent toutes deux la forme diagonale. Deux matrices diagonales étant toujours permutables, la proposition est établie.

La condition est nécessaire. Supposons donc que nous ayons (dans un système quelconque de coordonnées)

$$AB = BA$$

et supposons que, par un changement convenable de coordonnées, nous ayons ramené la matrice A à la forme diagonale [1] $A = \| \delta_{ik} \lambda_i \|$. La matrice B prendra alors la forme $B = \| b_{ik} \|$. En écrivant que ces deux

---

[1] Cette forme diagonale se compose de la succession, le long de la diagonale principale, des matrices $A_1, \ldots, A_p$ correspondant aux valeurs caractéristiques distinctes $\lambda^{(1)}, \ldots, \lambda^{(p)}$ de A. On a

$$A_i = \left\| \begin{matrix} \lambda^{(i)} & 0 & . & 0 \\ 0 & \lambda^{(i)} & . & 0 \\ & & \ddots & \\ 0 & . & 0 & \lambda^{(i)} \end{matrix} \right\| \updownarrow s_i.$$

matrices sont permutables, il vient

$$\lambda_i b_{ik} = b_{ik} \lambda_k,$$

soit

$$b_{ik}(\lambda_i - \lambda_k) = 0.$$

Si $\lambda_i \neq \lambda_k$, cette relation entraîne $b_{ik} = 0$. Considérons donc les valeurs caractéristiques distinctes $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p)}$ auxquelles correspondent les multiplicités $E^{(1)}, E^{(2)}, \ldots, E^{(p)}$. Les éléments b_{ik} seront donc nuls lorsque les indices i et k correspondront à deux multiplicités différentes. Il en résulte que la matrice B se décompose en une succession de p matrices partielles d'ordres $s_1, s_2, \ldots, s_p$, en désignant par ces nombres les ordres de multiplicité des valeurs caractéristiques $\lambda^{(i)}$, c'est-à-dire les nombres de dimensions des sous-espaces $E^{(1)}, E^{(2)}, \ldots, E^{(p)}$, les matrices B_i correspondant, lignes pour lignes et colonnes pour colonnes, aux matrices A_i de la réduite A

$$B = \begin{Vmatrix} B_1 & & & 0 \\ & B_2 & & \\ & & \ddots & \\ 0 & & & B_p \end{Vmatrix}.$$

Les matrices partielles $B_1, B_2, \ldots, B_p$ sont hermitiennes comme B; or nous savons que la matrice A conservera la forme canonique si nous effectuons dans chacun des sous-espaces $E^{(1)}, E^{(2)}, \ldots, E^{(p)}$ un *changement unitaire quelconque de coordonnées*. Nous choisirons donc les axes de coordonnées situés dans le sous-espace $E^{(1)}$ tels que la matrice hermitienne B_1 se réduise à la forme diagonale; et nous ferons de même pour les autres sous-espaces. Finalement nous sommes parvenus à *réduire simultanément* à la forme diagonale les deux matrices hermitiennes permutables : le théorème est démontré.

Ce théorème peut s'étendre *à un nombre quelconque fini ou à une infinité d'opérateurs* :

La condition nécessaire et suffisante pour qu'un nombre fini ou une infinité d'opérateurs (ou matrices) d'Hermite soient deux à deux permutables est qu'une même substitution unitaire les réduise tous à la forme diogonale (1).

(1) *Voir* WEYL, Chap. I, § 5.

7. Décomposition spectrale d'un opérateur hermitien A. — Nous exposerons cette question sur les opérateurs ou les formes d'Hermite, les notations étant plus simples. Hilbert, puis Hellinger ont employé une méthode de réduction, qui a permis d'en faire l'*extension à des classes très générales d'opérateurs* hermitiens de l'espace à une infinité de dimensions, où le spectre présente une infinité de valeurs qui ne sont pas nécessairement discontinues.

Considérons un opérateur A et ses valeurs propres distinctes rangées par ordre croissant $\lambda^{(1)} < \lambda^{(2)} < \ldots < \lambda^{(p)}$ dont les ordres respectifs sont $s_1, s_2, \ldots, s_p$. Nous avons fait correspondre à chaque nombre réel α un sous-espace $E^{(\alpha)}$ formé par les vecteurs X satisfaisant à $(\alpha - A)X = 0$ et un opérateur hermitien E_α définissant la projection orthogonale sur le sous-espace $E^{(\alpha)}$. Rappelons que si α n'est pas égal à une valeur caractéristique, E_α est nul, tandis que si α est égal à une valeur caractéristique, E_α n'est pas nul, et que nous avons les relations

$$\begin{aligned} E_\alpha E_\alpha &= E_\alpha, \\ E_\alpha E_\beta &= 0 \quad \text{si} \quad \alpha \neq \beta, \end{aligned}$$

la première relation exprimant que les opérateurs sont des opérateurs de projection, la seconde qu'ils sont indépendants.

Hilbert introduit un opérateur (ou matrice) dépendant d'un paramètre λ, défini par la relation suivante :

$$E(\lambda) = \sum_{\alpha < \lambda} E_\alpha.$$

Par conséquent, lorsque λ varie, l'opérateur $E(\lambda)$ prend les valeurs

$$\begin{aligned} &\lambda \leqq \lambda^{(1)}, && E(\lambda) = 0, \\ &\lambda^{(1)} < \lambda \leqq \lambda^{(2)}, && E(\lambda) = E_{\lambda^{(1)}}, \\ &\ldots\ldots\ldots\ldots, && \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ &\lambda^{(2)} < \lambda \leqq \lambda^{(3)}, && E(\lambda) = E_{\lambda^{(1)}} + E_{\lambda^{(2)}}, \\ &\lambda^{(p)} < \lambda, && E(\lambda) = E_{\lambda^{(1)}} + E_{\lambda^{(2)}} + \ldots + E_{\lambda^{(p)}} = 1. \end{aligned}$$

Parallèlement nous introduirons

$$A(\lambda) = \sum_{\alpha < \lambda} \alpha E_\alpha.$$

Soit

$$\begin{aligned} &\lambda \leqq \lambda^{(1)}, && A(\lambda) = 0, \\ &\lambda^{(1)} < \lambda \leqq \lambda^{(2)}, && A(\lambda) = \lambda^{(1)} E_{\lambda^{(1)}}, \\ &\lambda^{(2)} < \lambda \leqq \lambda^{(3)}, && A(\lambda) = \lambda^{(1)} E_{\lambda^{(1)}} + \lambda^{(2)} E_{\lambda^{(2)}}, \\ &\ldots\ldots\ldots\ldots, && \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ &\lambda^{(p)} < \lambda, && A(\lambda) = \lambda^{(1)} E_{\lambda^{(1)}} + \lambda^{(2)} E_{\lambda^{(2)}} + \ldots + \lambda^{(p)} E_{\lambda^{(p)}} = A. \end{aligned}$$

Si les valeurs caractéristiques formaient un spectre continu, les expressions $E(\lambda)$ et $A(\lambda)$ varieraient continûment.

Donnons quelques propriétés des opérateurs $E(\lambda)$.

a. Nous avons

$$E(\lambda)E(\lambda) = E(\lambda),$$

car le premier membre s'écrit

$$\left(\sum_{\alpha<\lambda} E_\alpha\right)\left(\sum_{\alpha<\lambda} E_\alpha\right) = \sum_{\alpha<\lambda} E_\alpha,$$

en exprimant que les E_α sont des opérateurs de projection indépendants.

b. Si $\lambda < \mu$, nous avons

$$E(\lambda)E(\mu) = E(\mu)E(\lambda) = E(\lambda),$$

car dans ces conditions

$$E(\mu) = E(\lambda) + \sum_{\lambda\leqq\alpha<\mu} E_\alpha$$

et

$$E(\lambda)E(\mu) = E(\lambda)E(\lambda) + E(\lambda)\sum_{\lambda\leqq\alpha<\mu} E_\alpha.$$

Or, d'après la propriété *a*,

$$E(\lambda)E(\lambda) = E(\lambda)$$

et, d'après l'indépendance des opérateurs,

$$E(\lambda)\left(\sum_{\lambda\leqq\alpha<\mu} E_\alpha\right) = 0.$$

Nous établirions de la même manière que

$$E(\mu)E(\lambda) = E(\lambda).$$

Ces deux propriétés peuvent se résumer dans la relation suivante :

Si $\lambda \leqq \mu$,

$$E(\lambda)E(\mu) = E(\mu)E(\lambda) = E(\lambda).$$

c. Si λ tend vers $-\infty$, $E(\lambda)$ tend vers *zéro* (dans le cas particulier étudié la limite étant déjà atteinte pour $\lambda^{(1)}$.

d. Si λ tend vers $+\infty$, $E(\lambda)$ tend vers *un* (dans le cas particulier étudié la limite étant déjà atteinte pour $\lambda^{(p)}$.

Si une famille d'opérateurs hermitiens jouit de ces propriétés elle est dite une *décomposition de l'unité*. Il résulte de cette définition que : *à un opérateur hermitien* A *correspond une décomposition de l'unité qui permet d'exprimer* A(λ).

Si $\lambda \leqq \mu$, nous poserons

$$E(\mu) - E(\lambda) = \Delta_\lambda^\mu E(\lambda) = \Delta_a E(\lambda),$$

a désignant l'intervalle défini par $\lambda \leqq \alpha < \mu$. La notation ab *représentant la partie commune aux deux intervalles* a *et* b, nous avons

$$\Delta_a E(\lambda) \times \Delta_b E(\lambda) = \Delta_{ab} E(\lambda),$$

comme on peut s'en assurer immédiatement en s'appuyant sur les propriétés des opérateurs E_α : opérateurs de projection indépendants. En particulier, si les intervalles a et b sont disjoints ou contigus, la relation précédente devient :

$$\Delta_a E(\lambda) \times \Delta_b E(\lambda) = 0. \qquad \text{(Relation d'indépendance.)}$$

Si a se réduit à $\lambda^{(i)}$ on a $\Delta_{\lambda^{(i)}} = E_{\lambda^{(i)}}$. Partout ailleurs $\Delta_x = 0$.

Hellinger, dans le *Journal de Crelle* (t. 136) a généralisé ces notions et introduit des solutions propres différentielles.

8. Avant de poursuivre, nous allons maintenant rappeler rapidement quelques notions relatives à *l'intégrale de Stieltjes* [1].

Considérons deux fonctions f et g définies dans un intervalle $\alpha\beta$, f continue et g à variation bornée. Divisons cet intervalle en intervalles partiels; désignons par $\Delta_i g$ la variation de la fonction g dans le $i^{\text{ième}}$ intervalle et par ξ_i un point de cet intervalle. Formons alors la somme

$$\Sigma f(\xi_i)\Delta_i g$$

étendue à tous les intervalles partiels. On démontre que lorsque le nombre des intervalles partiels augmente indéfiniment, leur longueur tendant vers zéro, cette somme tend vers une limite indépendante du choix des intervalles partiels, que l'on appelle intégrale de Stieltjes et que l'on désigne par la notation

$$\int_\alpha^\beta f dg.$$

---

(1) *Voir* Stieltjes, *Annales de Toulouse*, 1894, 1895.

Envisageons maintenant une distribution $m(\lambda)$ de masses réparties sur l'axe réel; soit une fonction $f(\lambda)$ continue. D'après la définition de l'intégrale de Stieltjes, nous aurons

$$\int_\alpha^\beta f(\lambda)\,d_\lambda m = \lim\left[\Sigma f(\lambda_i)\,\Delta_i m\right],$$

en désignant par $\Delta_i m$ la masse répartie sur le $i^{\text{ième}}$ intervalle partiel et par λ_i un point de cet intervalle. Dans le cas particulier où la distribution de masses a une dérivée (ou densité) en chaque point, l'intégrale de Stieltjes se ramène à une intégrale ordinaire

$$\int_\alpha^\beta f(\lambda)\,d_\lambda m = \int_\alpha^\beta f(\lambda)\frac{dm}{d\lambda}\,d\lambda.$$

Mais l'intégrale de Stieltjes peut aussi exister sans que la distribution de masses présente une densité; par exemple, supposons que la masse soit concentrée aux points $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p)}$ et désignons par m_i la masse située au point $\lambda^{(i)}$; nous aurons alors l'intégrale sous forme d'une somme de termes en nombre fini

$$\int_\alpha^\beta f(\lambda)\,d_\lambda m = \sum_i f(\lambda^{(i)})\,m_i.$$

9. Ce dernier exemple se rencontre en particulier dans le spectre discontinu que nous sommes en train d'étudier; les intégrales que nous allons écrire porteront sur des matrices, mais nous pouvons les considérer comme un langage symbolique indiquant les opérations à effectuer sur chaque terme des matrices. Nous aurons

$$\Delta_i E(\lambda) = 0,$$

si le $i^{\text{ième}}$ intervalle partiel ne contient pas de points du spectre;

$$\Delta_i E(\lambda) = E_{\lambda_j}$$

si le $i^{\text{ième}}$ intervalle partiel contient le point λ_j du spectre.

Par conséquent, les relations

$$1 = \Sigma E_\alpha,$$
$$A = \Sigma \alpha E_\alpha,$$

obtenues précédemment, peuvent s'écrire

$$1 = \int_{-\infty}^{+\infty} d_\lambda E(\lambda) = \lim\left[\Sigma \Delta_{a_i} E(\lambda)\right],$$

les a_i décomposition arbitraire de $(-\infty, +\infty)$

$$A = \int_{-\infty}^{+\infty} \lambda\, d_\lambda E(\lambda).$$

Cette forme intégrale est surtout intéressante parce qu'elle permet la généralisation de ces résultats.

Nous pouvons opérer *sur les formes* associées aux opérateurs ou aux matrices, au lieu d'opérer sur des matrices, ce qui donnera lieu à des *égalités entre nombres* (et non entre opérateurs) et à de *véritables intégrales de Stieltjes* (et non à des intégrales symboliques). Nous avons

$$E(\lambda, X, X) = X^* E(\lambda) X = [X, E(\lambda) X],$$
$$A(\lambda, X, X) = X^* A(\lambda) X = [X, A(\lambda) X],$$

$E(\lambda, X, X)$, est, pour chaque X, un nombre > 0 croissant avec λ.

Nous pouvons faire sur ces formes les mêmes raisonnements que nous avons faits sur les matrices. Nous obtiendrons les mêmes équations et nous aboutirons finalement aux deux relations

$$(X, X) = |X|^2 = \int_{-\infty}^{+\infty} d_\lambda E(\lambda, X, X),$$

$$A(X, X) = \int_{-\infty}^{+\infty} \lambda\, d_\lambda E(\lambda, X, X).$$

Citons encore la relation

$$(X, AY) = \int_{-\infty}^{+\infty} \lambda\, d_\lambda [X, E(\lambda) Y],$$

qui est équivalente à

$$A = \int_{-\infty}^{+\infty} \lambda\, d_\lambda E(\lambda)$$

et à

$$A(X, X) = \int_{-\infty}^{+\infty} \lambda\, d_\lambda E(\lambda, X, X).$$

Sous cette *forme intégrale*, Hilbert a montré que ces formules *sont encore vraies pour les opérateurs hermitiens bornés de l'espace de Hilbert.*

En résumé, un opérateur hermitien A fait correspondre à tout intervalle Δ, un opérateur hermitien de projection ΔE. Pour deux intervalles contigus Δ_1 et Δ_2, nous aurons

$$\Delta E = \Delta_1 E + \Delta_2 E,$$

en désignant par Δ l'intervalle formé par la réunion des précédents. Si les intervalles Δ_1 et Δ_2 sont sans point commun nous aurons

$$\Delta_1 E\, \Delta_2 E = 0.$$

et, de plus,

$$1 = \int_{-\infty}^{+\infty} d_\lambda E,$$

$$A = \int_{-\infty}^{+\infty} \lambda \, d_\lambda E.$$

10. La méthode de *décomposition spectrale* d'un opérateur hermitien peut être présentée sous une *forme matricielle purement algébrique* dont nous donnerons seulement ici un aperçu. (Pour le détail du calcul *voir*, par exemple, Wintner, *Spektraltheorie der unendlichen Matrizen*, p. 49.)

Soit $A = \| a_{ik} \|$ la matrice hermitienne envisagée. $U = \| u_{ik} \|$ la matrice unitaire réductrice, la matrice UAU^{-1} étant diagonale, $UAU^{-1} = \| \lambda_i \delta_{ik} \|$, les λ_i étant les valeurs propres de A. Cela veut dire que si $Y = AX$ est un opérateur de l'espace E_n, les vecteurs $\mathfrak{X} = UX$ et $\mathfrak{Y} = UY$ de l'espace transformé $\mathcal{E} = UE$ seront liés par l'opérateur

$$\mathfrak{Y} = UAU^{-1} \mathfrak{X}.$$

Considérons la matrice $E(\mu) = \| \sigma_{ik}(\mu) \|$ définie, *pour tout* μ *réel*, par

$$\sigma_{ik}(\mu) = \sum_{\lambda_j < \mu} \bar{u}_{ji} u_{jk}.$$

chaque $\sigma_{ik}(\mu)$ est une fonction discontinue de μ, variant par saut brusque lorsque μ traverse les valeurs propres *distinctes* $\lambda^{(1)}, \ldots, \lambda^{(p)}$ de A, constante dans l'intervalle de deux de ces valeurs propres.

On montre : 1° que, quel que soit l'ordre dans lequel les λ_j sont rangées, $E(\mu)$ *reste la même*; 2° que $E(\mu)$, constante dans l'intervalle de deux valeurs propres distinctes est telle que

$$E(\mu) = 0 \quad \text{si} \quad \mu \leqq \lambda^{(1)}, \text{ plus petite valeur propre,}$$

$$E(\mu) = 1 \quad \text{si} \quad \mu > \lambda^{(1)}, \text{ plus grande valeur propre,}$$

c'est-à-dire $E(-\infty) = 0$, $E(+\infty) = 1$, ce qui se vérifie par

$$\delta_{ik} = \int_{-\infty}^{+\infty} d\sigma_{ik}(\mu).$$

$E(\mu)$ est d'ailleurs hermitienne, quel que soit μ réel, et l'on vérifie que

$$a_{ik} = \int_{-\infty}^{+\infty} \mu \, d\sigma_{ik}(\mu).$$

3° On montre enfin que

$$E(\lambda) E(\mu) = E(\mu) E(\lambda) = E(\lambda) \quad \text{si} \quad \lambda \leqq \mu.$$

L'ensemble des propriétés précédentes montre que $E(\mu)$ possède toutes les propriétés étudiées aux n°s 7-9 et que l'on a bien

$$1=\int_{-\infty}^{+\infty} d_\mu E, \qquad A=\int_{-\infty}^{+\infty} \mu\, d_\mu E$$

ou encore

$$(X, AY)=\int_{-\infty}^{+\infty} \mu d_\mu (X, E(\mu)Y)$$

pour tout couple de vecteurs X, Y (¹).

II. — EXTENSION AUX OPÉRATEURS UNITAIRES.

1. Un opérateur unitaire U est caractérisé par la relation

$$UU^* = U^*U = 1.$$

En particulier, une matrice unitaire est *normale;* sa forme réduite est donc diagonale, soit

$$U = \| \lambda_i \delta_{ik} \|.$$

Nous avons alors

$$U^* = \| \bar{\lambda}_i \delta_{ik} \|$$

et

$$UU^* = \| \lambda_i \bar{\lambda}_i \delta_{ik} \| = 1.$$

Par conséquent

$$|\lambda_i|^2 = 1 \qquad \text{(quel que soit } i\text{)}.$$

Les *valeurs propres* d'une matrice unitaire ont donc pour *module* 1, autrement dit le *spectre est tout entier situé sur le cercle unité.*

Réciproquement on vérifie aussitôt qu'une matrice normale, dont le spectre est situé sur le cercle unité, est unitaire.

Étant donné un opérateur unitaire U, il est possible de trouver un système normal de coordonnées tel qu'en l'adoptant l'opérateur U soit

(¹) On vérifie aussi que *la résolvante de* A, c'est-à-dire la matrice

$$\| R_{ik}(\lambda) \| = (\lambda - A)^{-1}$$

est donnée par

$$R_{ik}(\lambda) = \int_{-\infty}^{+\infty} \frac{d\sigma_{ik}(\mu)}{\lambda-\mu}$$

ou, en abrégé,

$$(\lambda - A)^{-1} = \int_{-\infty}^{+\infty} \frac{d_\mu E}{\lambda-\mu}.$$

De même, on aura, Φ étant un *polynome quelconque à coefficients réels,*

$$\Phi(A) = \left\| \int_{-\infty}^{+\infty} \Phi(\mu)\, d\sigma_{ik}(\mu) \right\| = \int_{-\infty}^{+\infty} \Phi(\mu)\, d_\mu E.$$

ramené à la forme diagonale $U = \| \lambda_i \quad \delta_{ik} \|$. Ce que l'on peut exprimer par

$$U e_i = \lambda_i e_i,$$

e_i est appelé vecteur propre. Si λ_i est une racine caractéristique d'ordre s_j, il lui correspondra une multiplicité linéaire invariante $E^{(j)}$ à s_j dimensions, dans laquelle nous choisirons arbitrairement s_j vecteurs coordonnés rectangulaires. Les multiplicités $E^{(j)}$ sont orthogonales autrement dit les solutions propres correspondant à deux racines caractéristiques distinctes sont rectangulaires; démontrons-le directement.

Soient λ_1 et λ_2 deux racines caractéristiques distinctes et X_1, X_2 les solutions propres correspondantes. Nous avons

$$U X_1 = \lambda_1 X_1,$$
$$U X_2 = \lambda_2 X_2.$$

Multiplions la première relation par X_2; il vient

$$(U X_1, X_2) = \bar{\lambda}_1 (X_1, X_2).$$

Nous obtiendrons à partir de la seconde relation

$$X_2 = U^{-1} \lambda_2 X_2,$$

soit

$$X_2 = \lambda_2 U^{-1} X_2,$$

ou bien encore, puisque U est unitaire,

$$X_2 = \lambda_2 U^* X_2,$$

qui peut s'écrire

$$U^* X_2 = \lambda_2^{-1} X_2.$$

En tenant compte de cette relation, l'égalité

$$(U X_1, X_2) = X_1^* U^* X_2 = \bar{\lambda}_1 (X_1, X_2)$$

devient

$$\lambda_2^{-1} X_1^* X_2 = \bar{\lambda}_1 (X_1, X_2),$$

et en remarquant que λ_2 a 1 pour module, donc $\lambda_2^{-1} = \bar{\lambda}_2$,

$$\bar{\lambda}_2 (X_1, X_2) = \bar{\lambda}_1 (X_1, X_2).$$

Or λ_1 et λ_2 étant différents, il en est de même des nombres conjugués et, par conséquent

$$(X_1, X_2) = 0.$$

Remarquons que, dans chaque espace $E^{(i)}$ invariant par U, nous pouvons choisir un système normal arbitraire de coordonnées sans changer la forme diagonale. Il en résulte de même que, pour les matrices hermitiennes, le théorème suivant :

La condition nécessaire et suffisante pour qu'un nombre fini ou une infinité de matrices unitaires se réduisent par une même substitution unitaire à la forme diagonale, est qu'elles soient deux à deux permutables.

2. Relation de Cayley. — Il apparaît donc que les matrices unitaires et les matrices hermitiennes ont des propriétés voisines. Cela s'explique par la *relation de Cayley*.

Les valeurs caractéristiques d'un opérateur hermitien sont situées sur l'axe réel, tandis que celles d'un opérateur unitaire sont situées sur le cercle unitaire. Or on passe de l'axe réel au cercle unité par la transformation

$$z = \frac{x+i}{x-i}.$$

Nous allons donc essayer de former

$$U = \frac{H+i}{H-i},$$

H désignant un opérateur hermitien et nous chercherons à vérifier que U est alors un opérateur unitaire. Mais nous ne connaissons pas la division des matrices; par conséquent nous formerons

$$(H+i)(H-i)^{-1}.$$

Pour faire la démonstration nous allons *choisir des axes de coordonnées par rapport auxquels* H *prendra la matrice réduite*

$$H = \| h_j \quad \delta_{jk} \|,$$

les h_j étant tous réels, puisque H est hermitien. Nous aurons alors

$$H + i = \| (h_j + i) \quad \delta_{jk} \|,$$
$$H - i = \| (h_j - i) \quad \delta_{jk} \|.$$

Cette dernière matrice, n'ayant pas de valeur caractéristique nulle, présente une inverse

$$(H-i)^{-1} = \left\| \frac{\delta_{jk}}{h_j - i} \right\|$$

Le produit

$$U = (H+i)(H-i)^{-1}$$

est donc commutatif, puisque formé à l'aide de deux matrices diagonales et a pour expression

$$U = \left\| \frac{h_j + i}{h_j - i} \quad \delta_{ik} \right\|,$$

La matrice U est donc diagonale et a ses valeurs caractéristiques $\frac{h_j + i}{h_j - i}$ de module 1 ; elle est donc unitaire.

En résumé, *si la matrice* H *est hermitienne, la matrice*

$$U = (H + i)(H - i)^{-1}$$

est unitaire.

Réciproquement, la relation liant H et U peut se résoudre en H et s'écrire

$$H = i\,\frac{U + 1}{U - 1}.$$

Nous allons montrer que si U est une matrice unitaire, cette relation définit une matrice H hermitienne. En effet, nous pouvons prendre U sous forme réduite

$$U = \| \lambda_j \quad \delta_{jk} \| \qquad \text{avec} \quad |\lambda_j| = 1 \text{ quel que soit } j.$$

Nous aurons alors

$$U + 1 = \| (\lambda_j + 1) \quad \delta_{jk} \|,$$

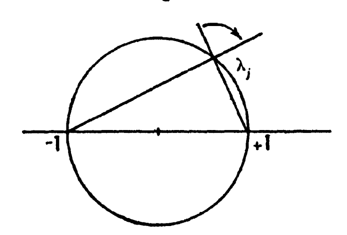
$$U - 1 = \| (\lambda_j - 1) \quad \delta_{jk} \|,$$

$$(U - 1)^{-1} = \left\| \frac{\delta_{jk}}{\lambda_j - 1} \right\|$$

et

$$H = i(U + 1)(U - 1)^{-1} = \left\| i\,\frac{\lambda_j + 1}{\lambda_j - 1} \quad \delta_{jk} \right\|.$$

Fig. 5.



Or λ_j étant sur le cercle unité, le nombre $\frac{\lambda_j + 1}{\lambda_j - 1}$ a pour argument $-\frac{\pi}{2}$, donc $\frac{\lambda_j + 1}{\lambda_j - 1}$ est réel. Il en résulte que l'opérateur H, dont la matrice se réduit à la forme diagonale, et dont les valeurs caractéristiques sont réelles est un opérateur hermitien.

Donc, les opérateurs unitaires et hermitiens se correspondent biuni-

voquement par les relations

$$U = \frac{H + i}{H - i},$$

$$H = i\frac{U + 1}{U - 1}$$

Ceci exprime le parallélisme entre les propriétés des opérateurs unitaires et hermitiens. Cette relation, introduite pour la première fois par Cayley, a été très utilisée par von Neumann; il l'a généralisée à l'espace de Hilbert, et cela lui a permis de traiter certains opérateurs hermitiens non bornés en les ramenant à des opérateurs unitaires bornés.

On doit à *Frobénius* l'extension suivante :

Si A est un opérateur normal non dégénéré, A^*A^{-1} est un opérateur unitaire.

Cette propriété se vérifie immédiatement en prenant la matrice A sous forme réduite

$$A = \| \lambda_i \quad \delta_{ik} \|;$$

dans ces conditions

$$A^* = \| \bar{\lambda}_i \quad \delta_{ik} \|,$$

$$A^{-1} = \left\| \frac{1}{\lambda_i} \quad \delta_{ik} \right\|$$

et alors

$$A^* A^{-1} = \left\| \frac{\bar{\lambda}_i}{\lambda_i} \quad \delta_{ik} \right\|.$$

Nous reconnaissons la forme diagonale de l'opérateur unitaire dont les valeurs caractéristiques ont 1 pour module.

3. **Les opérateurs unitaires infinitésimaux.** — Imaginons une rotation du corps des vecteurs. Dans le temps dt cette rotation transforme le vecteur X en le vecteur $X + dX$ et nous avons $\frac{dX}{dt} = CX$, C désignant une matrice ou un opérateur. Par abréviation nous dirons que C est un *opérateur unitaire infinitésimal*. Dans la transformation, les longueurs sont conservées; autrement dit $X^*X = \Sigma \bar{x}_i x_i$ est un invariant. Cela exige que

$$\sum \frac{d\bar{x}_i}{dt} x_i + \sum \bar{x}_i \frac{dx_i}{dt} = 0;$$

or, nous avons

$$\frac{dx_i}{dt} = \sum_k c_{ik} x_k.$$

La condition précédente peut donc s'écrire

$$\sum_i x_i \sum_k \overline{c_{ik}}\,\overline{x_k} + \sum_i \overline{x_i} \sum_k c_{ik} x_k = 0.$$

Soit encore, en changeant le nom des indices

$$\sum_{i,k} (c_{ik} + \overline{c_{ki}})\, \overline{x_i}\, x_k = 0.$$

Le premier membre de cette relation doit être nul quel que soit le vecteur X de composantes $x_1, x_2, \ldots, x_n$; pour cela il faut et il suffit que

$$c_{ik} + \overline{c_{ki}} = 0$$

ou

$$C^* = -C;$$

la matrice C doit présenter la *symétrie hermitienne gauche*. Elle n'est pas hermitienne, mais la matrice $H = \frac{C}{i}$ l'est. Nous aurons donc

$$\frac{dX}{dt} = iHX.$$

Si la matrice H *est constante et indépendante du temps*, nous en déduirons

$$X(t) = e^{iHt} X(0),$$

la matrice $U(t) = e^{iHt}$ étant unitaire, d'après son sens géométrique, et satisfaisant à la relation

$$U(t_1 + t_2) = U(t_1)\, U(t_2).$$

Si la matrice H n'est pas constante, nous effectuerons l'intégration de la relation différentielle par la méthode des approximations successives, en utilisant la forme

$$X(t) = X(0) + \int_0^t C(t) X(t)\, dt.$$

Nous aurons

$$
\begin{aligned}
X_0\ \ (t) &= X(0),\\
X_1\ \ (t) &= X(0) + \int_0^t C(t)\, X_0(t)\, dt,\\
X_2\ \ (t) &= X(0) + \int_0^t C(t)\, X_1(t)\, dt,\\
&\ldots\ldots\ldots\ldots\ldots\ldots,\\
X_{k+1}(t) &= X(0) + \int_0^t C(t)\, X_k(t)\, dt,\\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

Ce qui peut s'écrire

$$X_0(t) = X(o),$$

$$X_1(t) = X(o) + X(o)\int_{t_1=0}^{t_1=t} C(t_1)\,dt_1,$$

$$X_2(t) = X(o) + X(o)\int_{t_1=0}^{t_1=t} C(t_1)\,dt_1 + X(o)\int_{t_2=0}^{t_2=t}\int_{t_1=0}^{t_1=t_2} C(t_2)C(t_1)\,dt_1\,dt_2,$$

$$X_3(t) = X(o) + X(o)\int_{t_1=0}^{t_1=t} C(t_1)\,dt_1 + X(o)\int_{t_2=0}^{t_2=t}\int_{t_1=0}^{t_1=t_2} C(t_2)C(t_1)\,dt_1\,dt_2$$

$$+ X(o)\int_{t_3=0}^{t_3=t}\int_{t_2=0}^{t_2=t_3}\int_{t_1=0}^{t_1=t_2} C(t_3)C(t_2)C(t_1)\,dt_1\,dt_2\,dt_3,$$

....

et en posant

$$U_k(t) = \int\int\ldots\int C(t_k)C(t_{k-1})\ldots C(t_2)C(t_1)\,dt_1\,dt_2\ldots dt_k$$
$$(o \leqq t_1 \leqq t_2 \leqq \ldots \leqq t_k \leqq t),$$

nous aurons, pour terme général,

$$X_n(t) = X(o)\sum_{k=0}^{n} U_k(t).$$

Étant donné que

$$U_{k+1}(t) = \int_0^t C(t)\,U_k(t)\,dt,$$

il est facile de s'assurer que la fonction

$$X(t) = \lim_{n=\infty} X_n(t) = X(o)\sum_{k=0}^{\infty} U_k(t),$$

est bien solution du problème si la série ΣU_k est absolument et uniformément convergente dans l'intervalle où l'on opère, car il est alors possible de l'intégrer terme à terme. En particulier, si dans cet intervalle nous avons $|C(t)| < c$, il en résulte par un calcul très facile (*voir* métrique de Frobénius, Chap. II) que $|U_k(t)| < \frac{(ct)^k}{k!}$ et, par conséquent que la série ΣU_k est absolument et uniformément convergente dans un intervalle fini quelconque. Notons que si

$$U(t) = U(t, o) = \sum_{k=0}^{\infty} U_k(t),$$

on a

$$X(t_2) = U(t_2, t_1)X(t_1),$$
$$U(t_3, t_1) = U(t_3, t_2)U(t_2, t_1).$$

4. **Composantes hermitiennes d'une matrice quelconque.** — Soient A une matrice quelconque et A^* son associée. Envisageons la matrice $B = A + A^*$; nous avons $B^* = A^* + A = B$; la matrice B est donc hermitienne. Envisageons de même la matrice $C = A - A^*$; nous avons $C^* = A^* - A = -C = i^2 C$; il en résulte que $(iC)^* = iC$, donc que *la matrice* iC *est hermitienne* ainsi que $-iC = \frac{C}{i}$. Posons

$$A + A^* = 2A_1,$$
$$A - A^* = 2iA_2,$$

les matrices A_1 et A_2 sont hermitiennes et parfaitement déterminées et nous déduisons de ces deux relations

$$A = A_1 + iA_2,$$
$$A^* = A_1 - iA_2.$$

Les matrices A_1 et A_2 sont dites les composantes hermitiennes (uniques) de la matrice A; elles sont à rapprocher des composantes réelle et imaginaire d'un nombre complexe, ainsi qu'il apparaît par la considération de la forme associée à la matrice A.

$$X^*AX = X^*A_1X + iX^*A_2X,$$

les formes associées à A_1 et A_2 sont donc les composantes réelle et imaginaire de la forme associée à A, car les deux formes hermitiennes X^*A_1X et X^*A_2X sont réelles.

Démontrons maintenant que la condition nécessaire et suffisante pour qu'une matrice soit *normale est que ses composantes hermitiennes soient permutables*. En effet

$$AA^* = A_1^2 + iA_2A_1 - iA_1A_2 + A_2^2,$$
$$A^*A = A_1^2 + iA_1A_2 - iA_2A_1 + A_2^2.$$

Si nous exprimons que $AA^* = A^*A$, il vient immédiatement

$$A_1A_2 = A_2A_1.$$

III. — QUELQUES COMPLÉMENTS.

1. **Loi d'inertie.** — La réduction d'une forme d'Hermite en une somme de normes

$$A(X, X) = \sum_\alpha \alpha E_\alpha(X, X) \qquad [\alpha = \lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(p)}; \; E_i(X, X) = (E_iX, E_iX)]$$

peut être obtenue sans changement de coordonnées normales (mais avec des coordonnées obliques) par application du théorème classique de la décomposition en carrés des formes d'Hermite (et des formes quadratiques réelles).

Si une norme est affectée d'un coefficient positif, on peut faire entrer celui-ci dans la norme et n'avoir plus que le coefficient $+1$. Si une norme est affectée d'un coefficient négatif, on peut faire entrer la valeur absolue de celui-ci dans la norme et n'avoir plus que le coefficient -1. De telle sorte qu'une réduction de la forme d'Hermite $H(X, X)$ se présentera sous la forme

$$H(X, X) = \bar{y}_1 y_1 + \bar{y}_2 y_2 + \ldots + \bar{y}_r y_r - \bar{y}_{r+1} y_{r+1} - \bar{y}_{r+2} y_{r+2} - \ldots - \bar{y}_n y_n.$$

Elle comprendra r signes $+$ et $n-r$ signes $-$. Nous allons démontrer que, quelle que soit la décomposition de la forme d'Hermite, elle présentera toujours le même nombre de signes $+$ et de signes $-$; ce théorème correspond à la propriété bien connue de Sylvester pour les formes quadratiques réelles. Soit donc une autre décomposition

$$H(X, X) = \bar{z}_1 z_1 + \bar{z}_2 z_2 + \ldots + \bar{z}_s z_s - \bar{z}_{s+1} z_{s+1} - \bar{z}_{s+2} z_{s+2} - \ldots - \bar{z}_n z_n,$$

les z étant, de même que les y, des formes linéaires en x. Prouvons que nous avons nécessairement $r = s$. En effet

$$H(X, X) = \sum_1^r |y_i|^2 - \sum_{r+1}^n |y_i|^2 = \sum_1^s |z_i|^2 - \sum_{s+1}^n |z_i|^2,$$

Soit

$$\sum_1^r |y_i|^2 + \sum_{s+1}^n |z_i|^2 = \sum_1^s |z_i|^2 + \sum_{r+1}^n |y_i|^2.$$

Or si nous supposons que $r < s$, il en résulte que le premier membre est la somme de $r + n - s < n$ carrés; d'après la théorie des équations linéaires, il existe donc une solution non nulle au moins qui annule le premier membre en satisfaisant aux équations

$$\begin{aligned} y_1 &= 0, & y_2 &= 0, & &\ldots, & y_r &= 0, \\ z_{s+1} &= 0, & z_{s+2} &= 0, & &\ldots, & z_n &= 0. \end{aligned}$$

Cette solution annule donc aussi le second membre; et puisque celui-ci se présente sous forme d'une somme de carrés, elle annule chaque terme, soit

$$\begin{aligned} y_{r+1} &= 0, & y_{r+2} &= 0, & &\ldots, & y_n &= 0, \\ z_1 &= 0, & z_2 &= 0, & &\ldots, & z_s &= 0. \end{aligned}$$

Il existerait donc une solution non nulle qui annulerait simultanément toutes les formes linéaires y_i; or ceci est impossible puisque ces formes sont au nombre de n, dépendent de n variables et sont linéairement indépendantes. Nous ne pouvons donc avoir $r < s$; pour la même raison, nous ne pouvons avoir $r > s$; donc $r = s$.

2. Remarques relatives à la résolvante d'Hilbert. — La matrice hermitienne H a pour résolvante d'Hilbert

$$R_\mu = \frac{(1-\mu H)^{-1} - 1}{\mu}.$$

Si nous effectuons dans R_μ le changement unitaire de variables

$$X = UY, \qquad x_i = \sum_k u_{ik} y_k,$$

qui permet de passer de $H(X, X)$ à sa forme réduite $K(Y, Y)$, nous obtenons

$$\frac{(1-\mu K)^{-1} - 1}{\mu};$$

c'est-à-dire la résolvante de $K(Y, Y)$.

Si, d'après les notations de Hilbert, nous posons $\mu_p = \frac{1}{\lambda_p}$ en désignant par λ_p les valeurs caractéristiques de la matrice hermitienne, nous voyons que la forme associée à $1 - \mu K$ est

$$\sum_p \bar{y}_p y_p - \mu \sum_p \frac{\bar{y}_p y_p}{\mu_p} = \sum_p \frac{\mu_p - \mu}{\mu_p} \bar{y}_p y_p.$$

Par conséquent la forme associée à $(1-\mu K)^{-1}$ est

$$\sum_p \frac{\mu_p}{\mu_p - \mu} \bar{y}_p y_p$$

et à $(1-\mu K)^{-1} - 1$ est

$$\sum_p \left(\frac{\mu_p}{\mu_p - \mu} - 1\right) \bar{y}_p y_p = \sum_p \frac{\mu}{\mu_p - \mu} \bar{y}_p y_p.$$

La forme associée à la résolvante de $K(Y, Y)$ est donc

$$\sum_{p=1}^{n} \frac{1}{\mu_p - \mu} \bar{y}_p y_p.$$

Et il nous est maintenant facile de revenir aux premières variables. La réduction est, en effet, obtenue par le changement *unitaire* de variables

$$Y = U^{-1}X = U^{*}X,$$

soit, en explicitant,

$$y_p = \sum_{i=1}^{n} \bar{u}_{ip} x_i \qquad (p = 1, 2, \ldots, n),$$

d'où la forme résolvante de H(X, X) sous l'expression remarquable d'une décomposition en carrés

$$X^{*}R_{\mu}X = \sum_{p=1}^{n} \frac{|u_{1p}\bar{x}_1 + u_{2p}\bar{x}_2 + \ldots + u_{np}\bar{x}_n|^2}{\mu_p - \mu}.$$

CHAPITRE V.

VALEURS D'UNE FORME ASSOCIÉE A UN OPÉRATEUR. LIEN AVEC LA RÉDUCTION CANONIQUE D'UNE FORME HERMITIENNE (MÉTHODE D'EXTREMUM POUR LA RÉDUCTION). DOMAINE DES VALEURS D'UNE FORME ASSOCIÉE A UN OPÉRATEUR QUELCONQUE.

I. — MÉTHODE D'EXTREMUM POUR LA RÉDUCTION DES FORMES D'HERMITE.

1. Les méthodes d'extremum pour l'établissement de théorèmes d'existence ont une grande simplicité; elles sont d'ailleurs fort anciennes puisqu'elles ont été utilisées par Legendre, Lagrange, Gauss, Dirichlet, Riemann, Hilbert, etc. Elles ont été appliquées systématiquement par Courant dans ses recherches relatives aux problèmes de physique mathématique.

Ici la base de la méthode est le *lemme de Bolzano-Weierstrass* dont on tire comme conséquence que : toute fonction réelle *continue* de K variables réelles dans un domaine borné et fermé est *uniformément continue*, donc elle est *bornée*, donc elle a un *maximum;* elle *atteint effectivement ce maximum*, pour un point au moins, du domaine des variables ou de sa frontière.

Ces propriétés classiques étant rappelées, nous allons donner le *premier état de la méthode.*

Envisageons la forme d'Hermite

$$H(X, X) = X^* HX = \Sigma h_{ik} \bar{x}_i x_k$$

et nous allons considérer les *valeurs prises par la forme* H, *lorsque* X *est un vecteur quelconque satisfaisant à* $|X|^2 = 1$. Le domaine de variation des x_i est borné et fermé. $H(X, X)$ est fonction réelle et continue des x_i dans ce domaine. La fonction H présente donc un maximum λ_1, atteint pour $X = e_1$. Considérons alors les vecteurs X de longueurs 1 *orthogonaux à* e_1, c'est-à-dire tels que $(e_1, X) = 0$; les variables X

décrivent encore un domaine borné et fermé dans lequel la fonction H atteint son maximum $\lambda_2 \leq \lambda_1$ pour $X = e_2$. Puis nous considérerons les vecteurs X orthogonaux à e_1 et à e_2; ils décrivent encore un domaine borné et fermé dans lequel la fonction H atteint son maximum $\lambda_3 \leq \lambda_2$ pour $X = e_3$, ..., et ainsi de suite. Les valeurs λ_i sont réelles comme les valeurs de H(X, X).

Nous avons ainsi mis en évidence un système orthonormal de vecteurs $e_1, e_2, \ldots, e_n$. Formons maintenant la matrice U dans laquelle les éléments d'une colonne sont les composantes de l'un de ces vecteurs, soit

$$U = \| e_1 \quad e_2 \quad \ldots \quad e_n \|,$$

et effectuons le changement de variables défini par

$$X = UY,$$

ce qui revient à prendre pour nouveaux vecteurs coordonnés les vecteurs de $e_1, e_2, \ldots, e_n$. La forme deviendra, en nouvelles variables,

$$Y^* U^* H U Y = Y^* K Y = K(Y, Y) = \sum_{ij} k_{ij} \bar{y}_i y_j,$$

c'est la forme associée à la matrice transformée. Montrons que c'est la forme canonique diagonale.

En effet, elle est maxima pour $X = e_1$, c'est-à-dire pour $Y^{(1)}$ de coordonnées nouvelles (1, 0, 0, ..., 0), et a pour valeur λ_1. Donc $k_{11} = \lambda_1$. Pour tous les vecteurs $y_1 = 0$ (orthogonaux à e_1) elle est maxima pour $X = e_2$, c'est-à-dire pour $Y^{(2)}$ de coordonnées nouvelles (0, 1, 0, 0, ..., 0), et a pour valeur λ_2, et ainsi de suite.

Posons

$$K(Y, Y) = \Sigma k_{ij} \bar{y}_i y_j = \lambda_1 \bar{y}_1 y_1 + \ldots$$

et formons la différence,

$$K(Y, Y) - \lambda_1 (\bar{y}_1 y_1 + \bar{y}_2 y_2 + \ldots + \bar{y}_n y_n) = K_1(Y, Y).$$

L'expression $K_1(Y, Y)$ ne contient plus de terme $\bar{y}_1 y_1$; d'autre part, elle est toujours *négative ou nulle* puisqu'elle l'est d'après la définition même de λ_1, lorsque $|Y| = 1$. Montrons que $K_1(Y, Y)$ ne contient pas la variable y_1; pour cela, faisons

$$y_1 = 1, \qquad y_2 = \varepsilon, \qquad y_3 = y_4 = \ldots = y_n = 0;$$

l'expression se réduit alors à

$$K_1(Y, Y) = \alpha_{12}\varepsilon + \alpha_{21}\bar{\varepsilon} + \alpha_{22}\varepsilon\bar{\varepsilon} \qquad (\alpha_{21} = \bar{\alpha}_{12} \text{ et } \alpha_{22} \text{ réel}).$$

Si ε est réel, la partie principale est

$$(\alpha_{12} + \alpha_{21})\varepsilon$$

et change de signe lorsque ε change de signe; l'hypothèse que $K_1(Y, Y)$ est toujours négative entraîne donc

$$\alpha_{12} + \alpha_{21} = 0.$$

Si ε est imaginaire pure $\varepsilon = i\varepsilon'$, la partie principale est

$$(\alpha_{12} - \alpha_{21})\varepsilon$$

et change de signe lorsque ε' change de signe (ne pas oublier que $\alpha_{12} - \alpha_{21}$ est imaginaire pur); l'hypothèse que $K_1(Y, Y)$ est toujours négative entraîne donc $\alpha_{12} - \alpha_{21} = 0$.

Nous avons donc nécessairement

$$\alpha_{12} = \alpha_{21} = 0;$$

autrement dit il n'y a pas de termes $y_1\bar{y}_2$ ni $y_1\bar{y}_2$. Il est clair que la même démonstration nous permet d'affirmer qu'il n'y a pas non plus de termes $\bar{y}_1 y_3$ ni $\bar{y}_3 y_1$, ..., $\bar{y}_1 y_n$ ni $\bar{y}_n y_1$.

Finalement nous voyons que $K_1(Y, Y)$ ne contient que les variables $y_2, \ldots, y_n$; il en est évidemment de même de l'expression

$$K(Y, Y) - \lambda_1 \bar{y}_1 y_1;$$

car

$$K(Y, Y) - \lambda_1 \bar{y}_1 y_1 = K_1(Y, Y) + \lambda_1(\bar{y}_2 y_2 + \ldots + \bar{y}_n y_n),$$

que nous désignerons *dorénavant* par $K_1(Y, Y)$ pour ne pas compliquer les notations. Nous avons donc obtenu une forme

$$K_1(Y, Y) = K(Y, Y) - \lambda_1 \bar{y}_1 y_1$$

hermitienne et ne contenant que les variables $y_2, \ldots, y_n$; lorsque l'on fait $y_1 = 0$, il est clair que la forme K se réduit à la forme K_1; or nous savons que, dans ces conditions, elle présente un maximum λ_2 obtenu pour $X = e_2$, c'est-à-dire $Y^{(2)}(0, 1, 0, 0, \ldots, 0)$; il en est donc de même de la forme K_1 sur laquelle le raisonnement précédent s'applique; nous obtenons donc une forme

$$K_2(Y, Y) = K_1(Y, Y) - \lambda_2 \bar{y}_2 y_2$$

hermitienne et ne contenant que les variables $y_3, \ldots, y_n$. Et nous pouvons ainsi continuer jusqu'à épuisement de toutes les variables; finalement nous voyons que la transformation $X = UY$ a ramené la forme hermitienne à son expression canonique

$$H(X, X) = K(Y, Y) = \sum_{i=2}^{n} \lambda_i \bar{y}_i y_i.$$

Cette méthode de recherche de la forme canonique nous a permis de donner une définition intéressante des valeurs caractéristiques successives

$$\lambda_1 \geqq \lambda_2 \geqq \lambda_3 \geqq \ldots \geqq \lambda_n.$$

Nous aurions pu rechercher les minima successifs au lieu des maxima successifs et obtenir les valeurs caractéristiques en commençant par la dernière; nous aurions été amené à faire les mêmes raisonnements.

Nous pouvions présenter la question d'une façon un peu différente. Il suffit, pour cela, de remarquer qu'il revient au même de rechercher les maxima successifs de $H(X, X)$ lorsque $|X| = 1$ ou de rechercher les maxima successifs de $\frac{H(X, X)}{(X, X)}$ sans imposer à $|X|$ de condition; ou bien encore de *rechercher les minima successifs de* (X, X) *lorsque* $H(X, X) = 1$; c'est là le second aspect que nous voulions donner au problème. Nous chercherons donc le minimum de (X, X) lorsque $H(X, X) = 1$; c'est $\frac{1}{\lambda_1}$ atteint pour $X = e_1$. Puis nous chercherons le le minimum de (X, X) lorsque, de plus, nous nous imposons la condition $(e_1, X) = 0$; c'est $\frac{1}{\lambda_2}$ atteint pour $X = e_2$, etc. Cette seconde façon de procéder peut être interprétée géométriquement, car $H(X, X) = 1$ est l'équation d'une quadrique centrée à l'origine; nous cherchons donc sur la quadrique des points dont la distance au centre est minimum.

Les deux manières précédentes d'aborder la question : *maximum de* H *sur* $(X, X) = 1$, ou *minimum de* (X, X) *sur* $H(X, X) = 1$, sont classiques en calcul des variations et correspondent à la *dualité des problèmes d'extremum lié.*

2. Montrons que les λ_i ainsi obtenus (maxima successifs de H) ne sont autres que les valeurs propres de la forme H.

En effet, les λ_i sont les racines de l'équation

$$(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n) = 0,$$

soit,

$$\begin{vmatrix} \lambda-\lambda_1 & 0 & \dots & 0 \\ 0 & \lambda-\lambda_2 & \dots & 0 \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ 0 & 0 & \dots & \lambda-\lambda_n \end{vmatrix} = 0.$$

Le premier membre est le déterminant de la forme

$$\lambda\sum_{p=1}^{n}\bar{y}_p y_p - \sum_{p-1}^{n}\lambda_p\bar{y}_p y_p = \lambda E(Y, Y) - K(Y, Y).$$

Or, nous savons que cette forme provient de

$$\lambda E(X, X) - H(X, X),$$

par le changement de variables

$$X = UY.$$

Il en résulte que le déterminant de $\lambda E(Y, Y) - C(Y, Y)$ est égal au déterminant de $\lambda E(X, X) - H(X, X)$; autrement dit, les λ_i sont les racines de l'équation

$$|\lambda\delta_{ik} - a_{ik}| = 0,$$

qui est l'*équation caractéristique de* H. Les λ_i sont donc les valeurs propres de H. Ce sont des nombres réels égaux aux *inverses des carrés des longueurs des demi-axes de la quadrique à centre d'équation*

$$H = 1.$$

La forme H est définie positive dans le cas où tous les λ_i sont positifs, et dans ce cas seulement. Elle ne dégénère que si un λ_i au moins est nul.

Les valeurs λ_i apparaissent bien comme des *valeurs privilégiées de la forme* H, ce qui justifie la dénomination de *valeurs propres ou valeurs caractéristiques de* H.

Les *itérées de* H s'obtiennent *immédiatement* sur la forme réduite. Si

$$H(X, X) = K(Y, Y) = \sum_{i}^{n}\lambda_p\bar{y}_p y_p,$$

on aura

$$H^l(X, X) = K^l(Y, Y) = \sum_{1}^{n}\lambda_p^l\bar{y}_p y_p$$

et l'on voit que les H^{2m} sont toutes définies positives si H n'est pas dégénérée.

3. La méthode que nous venons d'exposer a un très grave défaut. Le premier maximum λ_1 se déduit directement de la donnée de la forme; mais le second exige la détermination du premier, le troisième exige la détermination des deux premiers. En sorte que les λ_i *ne se déterminent pas individuellement et indépendamment*, mais *successivement*. Aussi Fischer (1) a-t-il donné un *deuxième état de la méthode* appelé : *méthode des plus petits maxima* dans lequel les λ peuvent être déterminés *indépendemment les uns des autres*. Courant a repris et considérablement développé l'idée de Fischer.

Considérons la forme hermitienne

$$A(X, X) = \Sigma a_{ik} \bar{x}_i x_k$$

et cherchons son maximum lorsque le vecteur X a une longueur égale à 1 et lorsqu'il est assujetti en outre à $h-1$ équations linéaires

$$(1) \qquad \sum_{q=1}^{n} \alpha_{pq} x_q = 0 \qquad \text{pour} \qquad p = 1, 2, \ldots, h-1.$$

Lorsque X balaie ce domaine borné et fermé, *la forme* A *présente un maximum* $\mathfrak{M}_h(\alpha_{pq})$ qui est fonction des coefficients constants α_{pq} déterminant le domaine. Si nous donnons à ces coefficients *toutes les valeurs possibles*, $\mathfrak{M}_h$ *a un minimum;* nous allons prouver que *c'est la* $h^{\text{ième}}$ *valeur caractéristique* λ_h.

A cet effet, nous allons utiliser la forme canonique

$$A(X, X) = K(Y, Y) = \lambda_1 \bar{y}_1 y_1 + \lambda_2 \bar{y}_2 y_2 + \ldots + \lambda_n \bar{y}_n y_n.$$

Avec les variables y_i, les h équations linéaires deviennent

$$(2) \qquad \sum_{q=1} \beta_{pq} y_q = 0 \qquad \text{pour} \qquad p = 1, 2, \ldots h-1.$$

Choisissons alors, en particulier,

$$y_{h+1} = y_{h+2} = \ldots = y_n = 0.$$

Les $h-1$ équations homogènes et linéaires, qui comprennent encore h inconnues $y_1, \ldots, y_h$, présentent, au moins, une solution non nulle et définie seulement à un paramètre multiplicatif près; nous pouvons déterminer celui-ci de façon que le vecteur solution ait une longueur égale à

(1) E. Fischer, *Monatshefte für Math. und Phys.*, 1905.

l'unité. Pour cette solution nous aurons

$$K(Y, Y) = \lambda_1 \bar{y}_1 y_1 + \lambda_2 \bar{y}_2 y_2 + \ldots + \lambda_h \bar{y}_h y_h \geqq \lambda_h [\bar{y}_1 y_1 + \ldots + \bar{y}_h y_h].$$

Or le vecteur solution ayant une longueur unité, il en résulte que

$$\bar{y}_1 y_1 + \bar{y}_2 y_2 + \ldots + \bar{y}_h y_h = 1, \qquad \text{car les} \qquad y_{h+k} = 0,$$

donc

$$K(Y, Y) \geqq \lambda_h.$$

Par conséquent le maximum de $K(Y, Y)$ est supérieur ou égal à λ_h, *quel que soit le choix des* β; donc le maximum de $A(X, X)$ est supérieur ou égal à λ_h, *quel que soit le choix des* α.

Or choisissons les valeurs de α dans (1) ou celles de β dans les équations (2), de telle sorte que les équations (2) se réduisent à

$$y_1 = 0,\ y_2 = 0,\ \ldots,\ y_{h-1} = 0 \qquad \text{(linéairement indépendantes)};$$

la forme se réduira alors à

$$K(Y, Y) = \lambda_h \bar{y}_h y_h + \lambda_{h+1} \bar{y}_{h+1} y_{h+1} + \ldots + \lambda_n \bar{y}_n y_n \leqq \lambda_h [\bar{y}_h y_h + \ldots + \bar{y}_n y_n].$$

Elle atteindra son maximum λ_h lorsque

$$y_{h+1} = y_{h+2} = \ldots = y_n = 0.$$

Le nombre λ_h *est donc un maximum et c'est le plus petit des maxima.* Le théorème est donc établi.

On peut choisir les α *de façon que le maximum de* $A(X, X)$ *atteigne sa plus petite valeur* λ_p.

En résumé, *si les valeurs propres de* A *sont rangées par ordre non croissant :* $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_n$, *la* $h^{\text{ième}}$ *valeur propre* λ_h *est la plus petite valeur que peut prendre le maximum de* $A(X, X)$ *lorsque* X, *satisfaisant à* $|X| = 1$, *est en outre assujetti à vérifier* $h - 1$ *équations linéaires arbitraires.* Cette plus petite valeur λ_h est atteinte lorsque ces $h - 1$ équations linéaires sont convenablement choisies (elles sont alors indépendantes) : si $X = UY$ (donc $Y = U^{-1}X = U^*X$) exprime la transformation unitaire réductrice de $A(X, X)$, les $h - 1$ équations précédentes s'expriment par $y_1 = 0, \ldots, y_{h-1} = 0$, où les y_i sont les formes linéaires en x données par $Y = U^{-1}X = U^*X$.

4. De cette deuxième méthode nous pourrions aussi déduire une *méthode corrélative* en fixant la valeur de la forme et en cherchant le

minimum de la longueur du vecteur. Donnons simplement un exemple. Soit un ellipsoïde et plaçons-nous dans le domaine réel

$$H(X, X) = \Sigma a_{ik} x_i x_k = 1.$$

Désignons par ξ_i les cosinus directeurs de la direction X; nous aurons alors, pour $i = 1, 2, \ldots, n$,

$$x_i = \rho\xi_i \qquad \text{avec} \qquad \Sigma\xi_i^2 = 1,$$

donc

$$\rho^2 \Sigma a_{ik} \xi_i \xi_k = 1.$$

Soit

$$\rho^2 = \frac{1}{H(\xi, \xi)},$$

ρ représente la longueur du vecteur *dont l'extrémité décrit l'ellipsoïde*.

Le carré du demi petit axe de l'ellipsoïde est l'inverse du maximum de $H(\xi, \xi)$ lorsque le vecteur ξ a pour longueur 1, soit

$$\rho_1^2 = \frac{1}{\lambda_1} \qquad (\text{ici } h = 1).$$

Imposons-nous maintenant une relation linéaire

$$\sum_i \alpha_i \xi_i = 0 \qquad (\text{ici } h = 2),$$

nous obtenons une section diamétrale. Cherchons alors le maximum de son petit axe : c'est l'axe moyen de l'ellipsoïde. Or le carré du demi petit axe de la section est l'inverse du maximum de $H(\xi, \xi)$ sur la section; la plus petite valeur de ces maxima est donc le carré du demi-axe moyen de l'ellipsoïde $\rho_2^2 = \frac{1}{\lambda_2}$.

Enfin, imposons-nous deux relations linéaires ($h = 3$); nous obtenons une direction et cherchons le maximum de la longueur découpée par l'ellipsoïde sur cette direction : c'est le grand axe de l'ellipsoïde. Il a un carré égal à l'inverse du minimum de $H(\xi, \xi)$ lorsque le vecteur ξ a pour longueur 1, soit

$$\rho_3^2 = \frac{1}{\lambda_3}.$$

5. **Applications.** — Cette définition des valeurs caractéristiques par les plus petits maxima est très importante. Elle permet d'étudier *directement les valeurs propres* sans passer par le système algébrique $AX = \lambda X$ qui les introduit en bloc sans en bien montrer les différences.

En particulier, cette méthode va nous permettre de résoudre très facilement le problème suivant :

Considérons la forme

$$H(X,X)=\sum_{i,k} h_{ik}\bar{x}_i x_k$$

et supposons que le vecteur X soit non seulement assujetti à rester sur la sphère $|X|=1$ mais, de plus, à satisfaire à p relations linéaires indépendantes

$$(3) \qquad \sum_{q=1}^{n} l_{mq}x_q=0 \qquad \text{pour} \qquad m=1, 2, \ldots p.$$

La forme H de n variables indépendantes devient donc une forme $\mathcal{H}$ ne dépendant plus que de $n-p$ *variables indépendantes seulement*. Nous nous proposons de *chercher les valeurs caractéristiques* Λ *de cette nouvelle forme*.

La $h^{\text{ième}}$ valeur propre Λ_h de $\mathcal{H}$ est le plus petit maximum de $\mathcal{H}$ lorsque nous ajoutons $h-1$ relations linéaires arbitraires, c'est-à-dire le plus petit maximum de H lorsqu'il y a $p+h$ relations linéaires dont h sont arbitraires, les p premières étant (3). Si toutes les relations étaient arbitraires, nous aurions un champ *plus étendu* de maxima et λ_{p+h}, qui en est le plus petit, satisfait donc à

$$\lambda_{p+h} \leqq \Lambda_h.$$

Si nous n'avions que les h relations arbitraires et si les autres relations (3) n'existaient pas, chaque maximum se trouverait augmenté et il en serait de même de leur borne inférieure Λ_h qui est devenue λ_h ; donc

$$\Lambda_h \leqq \lambda_h.$$

Finalement, nous avons encadré les nouvelles valeurs caractéristiques

$$\lambda_{p+h} \leqq \Lambda_h \leqq \lambda_h.$$

Si, dans une forme d'Hermite H *à n variables, on impose p relations indépendantes linéaires et homogènes de liaison à ces variables, elle devient une forme d'Hermite* $\mathcal{H}$ *à n — p variables. Les valeurs caractéristiques* Λ *de* $\mathcal{H}$ *satisfont aux inégalités suivantes :*

$$\begin{array}{ccccccc} \lambda_1 & \geqq & \lambda_2 & \geqq \ldots \geqq & \lambda_{n-p}, \\ \vee\!\shortparallel & & \vee\!\shortparallel & & \vee\!\shortparallel \\ \Lambda_1 & \geqq & \Lambda_2 & \geqq \ldots \geqq & \Lambda_{n-p}, \\ \vee\!\shortparallel & & \vee\!\shortparallel & & \vee\!\shortparallel \\ \lambda_{p+1} & \geqq & \lambda_{p+2} & \geqq \ldots \geqq & \lambda_n, \end{array}$$

où les λ *sont les valeurs caractéristiques de* H.

En particulier, si nous prenons $x_n = 0$ comme équation de liaison (3), la forme d'Hermite H se réduit à sa $(n-1)^{\text{ième}}$ *section*. La $h^{\text{ième}}$ valeur propre de la $(n-1)^{\text{ième}}$ section de H est comprise entre la $h^{\text{ième}}$ valeur propre de H et la $(h+1)^{\text{ième}}$. Et le même théorème est applicable de proche en proche pour les sections de tous les ordres de telle sorte que :

les valeurs caractéristiques de la $(n-1)^{\text{ième}}$ *section séparent celles de* H;

les valeurs caractéristiques de la $(n-2)^{\text{ième}}$ *section séparent celles de la* $(n-1)^{\text{ième}}$;

les valeurs caractéristiques de la $(n-3)^{\text{ième}}$ *section séparent celles de la* $(n-2)^{\text{ième}}$;

et ainsi de suite. Nous obtenons donc ici un théorème analogue au théorème de Rolle pour la distribution des zéros des dérivées successives d'une fonction.

Comme *autre application*, remarquons que si à une forme d'Hermite H(X, X) on ajoute une forme d'Hermite $H_1(X, X)$ non négative, les valeurs propres de $H + H_1$ seront supérieures ou égales aux valeurs propres correspondantes de H.

Pour les applications immédiates à la mécanique des petites oscillations on pourra consulter l'article de Courant intitulé *Zur Theorie der Kleinen Schwingungen* dans *Zeitschrift für angew. Math. und Mech.*, 1922, BD 2, S, 278-285.

II. — ÉTUDE PARTICULIÈRE DU DOMAINE DES VALEURS D'UNE FORME ASSOCIÉE A UN OPÉRATEUR QUELCONQUE.

1. A un opérateur A quelconque de matrice $\| a_{ik} \|$, nous associons la forme

$$A(X, X) = X^* AX = \Sigma a_{ik} \bar{x}_i x_k.$$

Nous allons étudier le *domaine* V(A) *du plan complexe* parcouru par le point A(X, X) lorsque le vecteur X décrit la sphère Σ, $|X| = 1$, ou $\Sigma |x_i|^2 = 1$. Le domaine de la variable X étant borné et fermé, la fonction A(X, X) étant continue, son module présente un maximum M(A), atteint en un point au moins du domaine et nous avons

$$|X^* AX| \leqq M(A).$$

Faisons maintenant une remarque essentielle pour la suite :

Si nous effectuons sur X une substitution unitaire quelconque définie par $X = UY$, la forme devient

$$X^* A X = Y^* U^* A U Y$$

et la variable Y, de même que la variable X, décrit le domaine $|Y| = 1$. Il en résulte que *le domaine* V(A) *est le même pour* A *et pour* $U^* A U = U^{-1} A U$; le domaine est *invariant* lorsqu'on effectue une transformation unitaire sur la matrice. Ce qui va nous permettre, dans l'étude qui va suivre, de *remplacer une matrice par sa forme canonique dans le groupe unitaire.*

Cas d'un opérateur hermitien. — En le ramenant à sa forme canonique diagonale, nous avons déjà vu que le domaine d'un opérateur hermitien est le *segment réel* (λ, Λ) *compris entre la plus petite et la plus grande de ses valeurs propres.*

Conséquence. — Si une *forme d'Hermite* A *est non négative*, nous avons

$$\Sigma a_{ik} \bar{x}_i x_k \geqq 0.$$

En faisant $x_i = 1$ et toutes les autres variables nulles, nous obtenons

$$\boxed{a_{ii} \geqq 0.}$$

En annulant toutes les variables, sauf x_i et x_k, nous obtenons

$$a_{ii} \bar{x}_i x_i + a_{ik} \bar{x}_i x_k + a_{ki} x_i \bar{x}_k + a_{kk} \bar{x}_k x_k \geqq 0.$$

Ceci exprime que la matrice hermitienne

$$\begin{Vmatrix} a_{ii} & a_{ik} \\ a_{ki} & a_{kk} \end{Vmatrix}$$

est non négative, c'est-à-dire que ses valeurs propres sont non négatives, donc que son déterminant est positif ou nul; nous obtenons ainsi une autre relation

$$\boxed{a_{ii} a_{kk} \geqq a_{ik} a_{ki} = |a_{ik}|^2.}$$

Cas d'un opérateur normal. — Nous le ramenons aussi à sa forme canonique $\sum_i \lambda_i \bar{x}_i x_i$, dans laquelle les λ_i sont des *nombres complexes quelconques*. Le problème consiste donc à chercher le domaine balayé

par l'expression $\Sigma\lambda_i\alpha_i$ lorsque les nombres α_i, tous positifs, satisfont d'autre part à la condition $\Sigma\alpha_i = 1$; autrement dit, à chercher le domaine balayé par le centre de gravité du système des points λ_i affectés de masses positives arbitraires. Nous savons que l'on démontre par récurrence, à partir de deux points λ_i, que ce domaine est le *plus petit polygone convexe contenant tous les points* λ_i (¹); les sommets du polygone sont nécessairement des valeurs caractéristiques, mais le polynome peut contenir à son intérieur d'autres valeurs caractéristiques.

En particulier si *l'opérateur est unitaire*, ses valeurs caractéristiques sont situées sur le cercle unité et le domaine est donc un *polygone convexe inscrit dans le cercle unité et dont les sommets sont les valeurs caractéristiques*. Il en résulte que les valeurs prises par la forme ont toutes un module inférieur ou égal à 1.

2. Cas général. — Les valeurs caractéristiques ne suffisant pas, dans le cas général, à déterminer la forme (²), nous pouvons prévoir comme probable que ces mêmes valeurs caractéristiques ne suffiront pas à déterminer le domaine comme dans les cas précédents. Nous allons, malgré cela, obtenir sur lui quelques renseignements.

Théorème de Hausdorff-Tœplitz. — Tœplitz a démontré que *la frontière de* V(A) *est une courbe convexe;* Hausdorff a démontré que V(A) *lui-même est tout l'intérieur de cette courbe convexe.*

En effet, la forme A peut être décomposée en ses deux composantes hermitiennes

$$A = A_1 + iA_2.$$

Nous pouvons toujours supposer, par une transformation unitaire [qui, nous l'avons déjà vu, n'altère pas le domaine V(A)], avoir ramené la forme A_1 a son expression canonique

$$A(X, X) = A_1(X, X) + iA_2(X, X).$$

Le point $A_1(X, X)$, décrit le segment (λ, Λ) de l'axe réel, d'après l'étude des formes hermitiennes. Considérons alors un nombre réel a_1, compris entre λ et Λ; il existe un ensemble de points de $\Sigma(|X| = 1)$ pour lesquels $A_1(X, X) = a_1$, montrons que cet ensemble est *connexe*, c'est-à-dire que deux points quelconques de cet ensemble peuvent être joints par un chemin dont tous les points appartiennent à l'ensemble.

(¹) *Voir* par exemple Caratheodory, *Rendiconti del Circolo Math. di Palermo*, 1911.
(²) Elles suffisent pour un opérateur normal.

Soient donc deux points de la sphère $\Sigma(|X|=1)$:

Y de coordonnées $y_1\, y_2, \ldots, y_n$ avec $y_i = |y_i|\, e^{\eta_i}$,
Z de coordonnées $z_1, z_2, \ldots, z_n$ avec $z_i = |z_i|\, e^{\zeta_i}$,

satisfaisant à $A_1(Y, Y) = A_1(Z, Z) = a_1$. Or

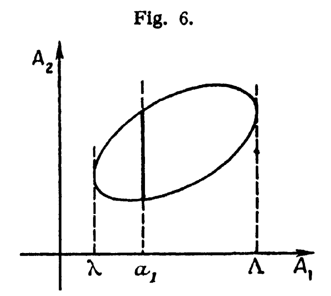
$$A_1(X, X) = \sum_i \lambda_i \bar{x}_i x_i;$$

les deux conditions précédentes s'écrivent donc

$$\sum_i \lambda_i |y_i|^2 = a_1,$$

$$\sum_i \lambda_i |z_i|^2 = a_1.$$

Fig. 6.



Envisageons maintenant les trois courbes suivantes définies par le paramètre réel t variant de o à 1 :

1° La courbe C_1, décrite par le point $X^{(1)}$ tel que $x_i^{(1)} = |y_i|\, e^{(1-t)\eta_i}$;
2° La courbe C_2, décrite par le point $X^{(2)}$ tel que

$$x_i^{(2)} = \sqrt{(1-t)|y_i|^2 + t|z_i|^2};$$

3° La courbe C_3, décrite par le point $X^{(3)}$ tel que $x_i^{(3)} = |z_i|\, e^{t.\zeta_i}$.

La courbe C_1, dont le point courant a des cordonnées de module constant, satisfaisant à $|X|^2 = |Y|^2 = 1$, joint le point Y au point de coordonnées $|y_i|$.

La courbe C_2, dont le point courant a des coordonnées réelles, satisfait bien elle aussi à l'équation de la sphère $|X|^2 = 1$; elle joint le point de coordonnées $|y_i|$ à celui de coordonnées $|z_i|$.

La courbe C_3, dont le point courant a des coordonnées de module constant, satisfait à l'équation de la sphère $|X|^2 = 1$; elle joint le point de coordonnées $|z_i|$ au point Z. Donc la courbe $C_1 + C_2 + C_3$ joint Y et Z sur la sphère Σ.

Il nous reste à vérifier que le long de ces trois courbes, nous avons la relation $A_1(X, X) = a_1$, pour avoir démontré que le domaine des points X de la sphère tels que $A_1(X, X) = a_1$ est connexe. Or ceci est immédiat :

$$A_1(X^{(1)}, X^{(1)}) = \sum_i \lambda_i \bar{x}_i^{(1)} x_i^{(1)} = \sum_i \lambda_i |y_i|^2 = A_1(Y, Y) = a_1,$$

$$A_1(X^{(2)}, X^{(2)}) = \sum_i \lambda_i |x_i^{(2)}|^2$$

$$= (1-t)\sum_i \lambda_i |y_i|^2 + t\sum_i \lambda_i |z_i|^2 = (1-t)A_1(Y, Y) + tA_1(Z, Z) = a_1,$$

$$A_1(X^{(3)}, X^{(3)}) = \sum_i \lambda_i |x_i^{(3)}|^2 = \sum_i \lambda_i |z_i|^2 = a_1.$$

Le domaine des points X envisagés est donc connexe; d'autre part, il est fermé puisque la fonction $A_1(X, X)$ qui sert à le définir est continue, et il est borné puisque $|X| = 1$. La forme $A_2(X, X)$ étant continue et réelle prend dans ce domaine toutes les valeurs comprises entre un minimum a'_2 et un maximum a''_2 (qui peuvent être confondus). Donc, le domaine V(A) est coupé par la parallèle à l'axe imaginaire, d'équation $A_1 = a_1$, suivant un segment $(a'_2 a''_2)$ lorsque a_1, appartient à l'intervalle (λ, Λ), et en zéro point si a_1 est extérieur à cet intervalle.

Cette propriété du domaine V(A) d'être coupé par les parallèles à une direction suivant un segment n'appartient pas seulement à la direction parallèle à l'axe imaginaire, *mais à toutes les directions*. En effet, la matrice $e^{i\theta}A$ a un domaine $V(e^{i\theta}A)$ qui se déduit de V(A) par rotation d'un angle θ autour de l'origine; ce nouveau domaine jouit de la propriété parallèlement à l'axe imaginaire, quel que soit θ; il en résulte que le domaine V(A) jouit de la propriété relativement à toutes les directions. *Le domaine* V(A) *est donc convexe.*

3. Montrons que *le domaine* V(A) *contient le spectre.* Soit, en effet, α une valeur propre; il existe un vecteur X, de longueur 1, satisfaisant à

$$\alpha X = AX,$$

c'est-à-dire

$$X^* \alpha X = X^* AX,$$

ou encore

$$\alpha = X^* A X = A(X, X),$$

puisque

$$X^* \alpha X = \alpha X^* X \quad \text{et} \quad X^* X = (X, X) = 1.$$

La forme *prend donc la valeur propre* α, *lorsque* X *a comme valeur la solution propre correspondante*. Autrement dit, les *valeurs propres de* A *appartiennent à* V(A). Il en résulte que les valeurs du spectre ont des modules inférieurs à M(A) et que les parties réelles des valeurs du spectre sont comprises entre le minimum et le maximum de A_1 et, par suite, entre $-$M(A) et $+$M(A); de même, les parties imaginaires des valeurs du spectre sont comprises entre le minimum et le maximum de A_2, donc entre $-$M(A) et $+$M(A).

Voilà tout ce que nous pouvons dire sur le domaine d'une matrice non normale si nous n'en connaissons que le spectre. Nous avions prévu que nos résultats ne seraient pas plus précis, car se donner le spectre d'une matrice, revient à se donner la diagonale principale de la forme réduite, les termes situés en dessous de cette diagonale étant arbitraires et influant sur le domaine.

III. — DOMAINE DES VALEURS DE LA FORME BILINÉAIRE ASSOCIÉE A UN OPÉRATEUR QUELCONQUE.

1. A un opérateur A quelconque, nous associons la forme bilinéaire définie par

$$X' A Y = \sum_{i,k} a_{ik} x_i y_k.$$

Nous allons chercher le domaine W(A) du plan complexe parcouru par cette expression lorsque les vecteurs X et Y décrivent les sphères $|X| = 1$, $|Y| = 1$.

Il est évident que *le domaine* W(A) *contient le domaine* V(A), puisque celui-ci est obtenu pour $Y = \overline{X}$.

Si le domaine W(A) *contient un point* W_0, *il contient tout le cercle centré à l'origine et passant par* W_0. En effet la valeur W_0 était atteinte pour un certain couple de vecteurs, soit X_0, Y_0; dans ces conditions, le couple X_0, $e^{i\theta} Y_0$ nous donnera la valeur $e^{i\theta} W_0$; il est clair qu'en faisant varier θ, nous obtiendrons tous les points du cercle.

Mais, comme dans le paragraphe précédent, le domaine W(A) est connexe comme le domaine de X et Y; il en résulte qu'il est formé

par une couronne circulaire centrée à l'origine. Nous allons montrer que ce domaine contient la valeur zéro, par conséquent qu'il se réduit à un cercle. En effet, Y étant fixé, l'équation $X'AY = 0$ admet au moins une solution X de module 1 (une équation homogène à n inconnues). Le domaine W(A) est donc un cercle de rayon $M_1(A)$ en désignant par $M_1(A)$ le maximum de $|X'AY|$ lorsque $|X| = |Y| = 1$. D'après une remarque faite au début de ce paragraphe, nous avons

$$M_1(A) \geqq M(A).$$

Cette inégalité va être complétée par le théorème de Hellinger-Tœplitz. Remarquons, en passant, que si A est unitaire, le domaine W(A) est constitué par le cercle unité, car $|X'AY| \leqq 1$ (1) et, d'autre part, les valeurs caractéristiques sont situées sur la circonférence unité.

2. Théorème de Hellinger-Tœplitz. — On a toujours $M_1(A) \leqq 2M(A)$.

Étudions d'abord le cas particulier où *la matrice* A *est normale* et établissons qu'alors $\underline{M_1(A) = M(A)}$. Les domaines V et W n'étant pas altérés par une transformation unitaire, nous pouvons ramener la matrice normale A à sa forme canonique; donc nous aurons

$$X^* AX = \Sigma \lambda_i \bar{x}_i x_i,$$
$$X'AY = \Sigma \lambda_i x_i y_i.$$

La forme X^*AX ayant pour domaine le plus petit polygone convexe contenant le spectre, nous avons

$$M(A) = \text{maximum de } |\lambda_i|,$$

atteint pour $X = e_i$, d'indice i correspondant au maximum de $|\lambda_i|$.

La forme $X'AY$ vérifie

$$|X'AY| = \left|\sum_i \lambda_i x_i y_i\right| \leqq \max|\lambda_i| \, \Sigma|x_i y_i| = M(A)\, \Sigma|x_i y_i|;$$

et, d'après une identité classique,

$$[\Sigma|x_i y_i|]^2 \leqq [\Sigma|x_i|^2][\Sigma|y_i|^2] = 1;$$

donc

$$|X'AY| \leqq M(A),$$

ce qui entraîne

$$M_1(A) \leqq M(A).$$

(1) D'après l'inégalité de Cauchy-Schwarz.

Comme, d'autre part, nous avons démontré l'inégalité

$$M_1(A) \geqq M(A),$$

il en résulte

$$\underline{M_1(A) = M(A)} \quad \underline{\text{pour A normale.}}$$

Abordons maintenant le *cas général*. Montrons que, pour une matrice quelconque, nous avons $M_1(A) \leqq 2M(A)$. A cet effet, décomposons la matrice en ses deux composantes hermitiennes

$$A = A_1 + iA_2.$$

Nous avons évidemment

$$M_1(A) \leqq M_1(A_1) + M_1(A_2);$$

car la valeur absolue d'une somme est inférieure ou égale à la somme des valeurs absolues de ses termes; et l'on passe immédiatement aux bornes supérieures. Or A_1 est une matrice hermitienne, donc normale; d'après le cas particulier nous avons donc

$$M_1(A_1) = M(A_1)$$

et de même

$$M_1(A_2) = M(A_2).$$

Nous en déduisons

$$M_1(A) \leqq M(A_1) + M(A_2).$$

Or, il est clair que

$$M(A_1) \leqq M(A)$$

et de même

$$M(A_2) \leqq M(A).$$

Donc, finalement,

$$M_1(A) \leqq 2M(A).$$

Remarquons que les inégalités obtenues

$$M(A) \leqq M_1(A) \leqq 2M(A)$$

ne peuvent être resserrées, car chacune des deux égalités est possible :

$$M(A) = M_1(A) \quad \text{pour les matrices normales,}$$

$$M_1(A) = 2M(A) \quad \text{par exemple pour } A = \begin{Vmatrix} 0 & 1 \\ 0 & 0 \end{Vmatrix}.$$

IV. — APPLICATION AUX NORMES MATRICIELLES.

1. Si H est une matrice hermitienne, nous désignerons par $n(H)$ la plus petite des valeurs caractéristiques λ_i, par $m(H)$ la plus grande et par $M(H)$ le maximum des deux quantités $|m(H)|$ et $|n(H)|$; remarquons que $M(H)$ a donc encore le sens que nous lui attribuions dans le paragraphe précédent :

$$n(H) \leqq \lambda_i \leqq m(H), \qquad M_1(H) = M(H) = \text{Max}\,[\,|m(H)|,\ |n(H)|\,].$$

Si A est une matrice non hermitienne et A^* son associée, nous avons déjà introduit les deux matrices hermitiennes

$$AA^* \quad \text{et} \quad A^*A$$

et les deux formes associées, appelées *normes matricielles*

$$X^*AA^*X = \sum_k \Big|\sum_i a_{ik}x_i\Big|^2,$$

$$X^*A^*AX = \sum_i \Big|\sum_k a_{ik}x_k\Big|^2.$$

Il résulte de ces expressions que les deux normes sont des formes hermitiennes définies non négatives et même définies positives si zéro ne fait pas partie du spectre, c'est-à-dire si A n'est pas dégénérée; car la condition nécessaire et suffisante pour que AA^* soit dégénérée est que A le soit $[\text{dét}\,AA^* = |\text{dét}\,A|^2]$.

D'après l'inégalité de Schwarz, nous avons, puisque $|X| = |Y| = 1$:

$$|X'AY|^2 = \Big|\sum_i x_i \overbrace{\Big(\sum_k a_{ik}y_k\Big)}^{\xi_i}\Big|^2$$

$$\leqq \Big(\sum_i |x_i|^2\Big)\Big(\sum_i \Big|\overbrace{\sum_k a_{ik}y_k}^{\xi_i}\Big|^2\Big) = \sum_i \Big|\sum_k a_{ik}y_k\Big|^2 = Y^*A^*AY.$$

Et, le vecteur Y étant déterminé, on peut choisir X de façon à avoir

l'égalité en prenant :

$$x_l = \frac{\sum_k a_{lk} y_k}{\sqrt{\sum_l \left| \sum_k a_{lk} y_k \right|^2}} = \frac{\xi_l}{|\xi_l|}.$$

Nous venons donc d'établir que

$$|X'AY|^2 \leqq Y^* A^* A Y,$$

l'égalité étant possible pour un choix de X *dépendant du choix de* Y. Il résulte de cette inégalité que le maximum $[M_1(A)]^2$ du premier terme ne peut être supérieur au maximum $M(A^*A)$ du second; choisissons alors Y de façon à rendre le second terme maximum, nous savons qu'il est possible de déterminer X de façon à obtenir l'égalité; il en résulte que

$$\underline{[M_1(A)]^2 = M(A^*A).}$$

Remarquons d'autre part que les matrices AA^* et A^*A étant équivalentes dans le groupe unitaire, ont même spectre; donc leurs formes d'Hermite, associées ont même maximum et même minimum,

$$m(AA^*) = m(A^*A),$$
$$n(AA^*) = n(A^*A),$$
$$V(AA^*) = V(A^*A),$$

ce qui entraîne que $M(A^*A) = M(AA^*)$ et finalement :

$$[M_1(A^*)]^2 = [M_1(A)]^2 = M(AA^*) = M(A^*A).$$

CHAPITRE VI.

ÉQUATIONS INTÉGRALES DE FREDHOLM A NOYAUX DÉGÉNÉRÉS (NOYAUX DE GOURSAT).

---

1. Avant de clore cette première Partie du cours, nous voudrions montrer, par un exemple particulier, la répercussion sur le domaine fonctionnel des propriétés de l'espace à n dimensions que nous avons étudiées.

Pour préciser, les propriétés des opérateurs linéaires que nous avons étudiées *analytiquement* sous l'aspect du *calcul des matrices* peuvent s'étudier sous l'*aspect fonctionnel* qui, une fonction $\varphi(s)$ étant donnée, lui fait correspondre par une *transformation linéaire* $\mathfrak{T}(\varphi)$ une fonction $f(s) = \mathfrak{T}(\varphi)$ de manière que

$$\mathfrak{T}(\varphi_1 + \varphi_2) = \mathfrak{T}(\varphi_1) + \mathfrak{T}(\varphi_2)$$

et

$$\mathfrak{T}(\lambda\varphi) = \lambda\,\mathfrak{T}(\varphi) \qquad (\lambda \text{ nombre quelconque}).$$

Nous particulariserons encore en envisageant la *résolution des équations linéaires*, ou le *calcul de la matrice inverse* d'une matrice donnée, ou la détermination de l'opérateur inverse d'un opérateur donné.

Le parallélisme se marquera nettement si nous choisissons $\mathfrak{T}(\varphi)$ sous la forme de Fredholm

$$\mathfrak{T}(\varphi) = \varphi(s) - \mu \int_a^b k(s, t)\,\varphi(t)\,dt,$$

et si nous cherchons à faire l'inversion de cette transformation en résolvant par rapport à φ l'équation

$$\varphi(s) - \mu \int_a^b k(s, t)\,\varphi(t)\,dt = f(s).$$

Pour *mettre en évidence* le parallélisme avec l'inversion des transformations linéaires de l'espace E_n à n dimensions, nous envisagerons

des noyaux du type considéré par M. Goursat ([1])

$$k(s, t) = \sum_{i=1}^{p} a_i(s) b_i(t).$$

Nous avons déjà, dans l'Introduction, dit quelques mots de l'équation intégrale de Fredholm de deuxième espèce

$$\varphi(s) - \mu \int_a^b k(s, t) \varphi(t) \, dt = f(s). \tag{1}$$

Nous supposons que les fonctions $k(s, t)$ et $f(s)$ sont connues et nous recherchons les fonctions φ continues et satisfaisant à cette équation. Nous avons insisté sur une propriété fondamentale de cette équation : sa linéarité; sa résolution est liée à celle de l'équation *homogène associée*

$$\varphi(s) - \mu \int_a^b k(s, t) \varphi(t) \, dt = 0. \tag{2}$$

Un rôle important sera aussi joué par l'*équation transposée*

$$\psi(s) - \mu \int_a^b k(t, s) \psi(t) \, dt = g(s), \tag{1'}$$

et par l'équation homogène correspondante

$$\psi(s) - \mu \int_a^b k(t, s) \psi(t) \, dt = 0. \tag{2'}$$

Fredholm a traité le problème pour des noyaux $k(s, t)$ très généraux. Nous allons simplement nous occuper ici des noyaux $k(s, t)$ *dégénérés*, c'est-à-dire tels que

$$k(s, t) = \sum_{i=1}^{p} a_i(s) \, b_i(t),$$

les fonctions a_i et b_i étant continues; nous supposerons aussi que la fonction f est continue et nous ne chercherons que les solutions φ continues. Le noyau dégénéré peut présenter plusieurs décompositions ; supposons que ce soit la plus courte que nous venons d'écrire. Il en résulte que *les fonctions $a_i(s)$ sont linéairement indépendantes*, car dans le cas contraire, l'une d'entre elles, a_p par exemple, serait une

([1]) *Annales de Toulouse*, 1908; *Bull. Soc. Math. de France*, 1907.

combinaison linéaire des autres, soit

$$a_p = \gamma_1 a_1 + \gamma_2 a_2 + \ldots + \gamma_{p-1} a_{p-1},$$

d'où

$$k(s, t) = \sum_{i=1}^{p-1} a_i(s)[b_i(t) + \gamma_i b_p(t)],$$

et nous pourrions obtenir la représentation du noyau avec $p - 1$ couples de fonctions, ce qui est contraire à l'hypothèse. Les fonctions a_i sont donc linéairement indépendantes, ainsi que les fonctions b_i.

L'équation de Fredholm (1) s'écrit

$$\varphi(s) - \mu \int_a^b \left(\sum_{i=1}^{p} a_i(s) b_i(t) \right) \varphi(t)\, dt = f(s),$$

ou bien

$$\varphi(s) - \mu \sum_{i=1}^{p} \left[a_i(s) \int_a^b b_i(t) \varphi(t)\, dt \right] = f(s).$$

Si nous posons alors

$$x_i = \int_a^b b_i(t) \varphi(t)\, dt = (b_i, \varphi),$$

produit intérieur des deux fonctions b_i et φ dans l'espace fonctionnel, l'équation devient

$$(3) \qquad \varphi(s) = f(s) + \mu \sum_{i=1}^{p} x_i a_i(s).$$

Il est clair que φ sera déterminé dès que les x_i le seront.
Reportons cette expression dans la relation définissant x_i

$$x_i = \int_a^b b_i(t) f(t)\, dt + \mu \sum_{k=1}^{p} x_k \int_a^b b_i(t) a_k(t)\, dt.$$

En posant

$$c_{ki} = \int_a^b a_k(t) b_i(t)\, dt = (a_k, b_i)$$

et

$$f_i = \int_a^b b_i(t) f(t)\, dt = (b_i, f),$$

cette équation s'écrit

$$(4) \qquad x_i = f_i + \mu \sum_{k=1}^{p} c_{ki} x_k,$$

et nous obtenons p équations semblables en donnant à i les valeurs 1, 2, ..., p.

Si φ est une solution de l'équation de Fredholm (1), il est clair que les $x_i = (b_i, \varphi)$ correspondants satisfont au système d'équations (4).

Remarquons que *le système* (4) *peut être formé uniquement à partir de* f *et de* $k(s, t)$. Réciproquement, supposons que nous ayons une solution $x_1, x_2, \ldots, x_p$ du système (4). La relation (3) nous permet d'en déduire par (3) une fonction $\varphi(s)$. Montrons que cette fonction est solution de l'équation de Fredholm (1); en effet,

$$\int_a^b k(s, t)\varphi(t)\,dt = \int_a^b \left(\sum_{i=1}^{p} a_i(s)\,b_i(t)\right)\varphi(t)\,dt = \sum_{i=1}^{p} a_i(s)\int_a^b b_i(t)\,\varphi(t)\,dt,$$

et, puisque φ provient de la relation (3),

$$\begin{aligned}\int_a^b k(s, t)\varphi(t)\,dt &= \sum_{i=1}^{p} a_i(s)\int_a^b b_i(t)\left[f(t) + \mu\sum_{i=1}^{p} x_k a_k(t)\right]dt \\ &= \sum_{i=1}^{p} a_i(s)\int_a^b b_i(t) f(t)\,dt + \sum_{i=1}^{p} a_i(s)\left\{\mu\sum_{i=1}^{p} x_k\int_a^b b_i(t)\,a_k(t)\,dt\right\} \\ &= \sum_{i=1}^{p} a_i(s) f_i + \sum_{i=1}^{p} a_i(s)\left\{\mu\sum_{k=1}^{p} c_{ki}x_k\right\} \\ &= \sum_{i=1}^{p} a_i(s)\left\{f_i + \mu\sum_{k=1}^{p} c_{ki}x_k\right\}.\end{aligned}$$

Or les x_k sont solutions du système (4); cette relation peut donc s'écrire

$$\int_a^b k(s, t)\varphi(t)\,dt = \sum_{i=1}^{p} a_i(s)x_i,$$

soit, en utilisant de nouveau l'équation (3),

$$\varphi(s) - f(s) = \mu\int_a^b k(s, t)\varphi(t)\,dt.$$

Par conséquent, la fonction φ ainsi obtenue est bien une solution de l'équation de Fredholm.

En définitive, l'équation (1) est équivalente à

$$(4)\qquad x_i - \mu \sum_{k=1}^{p} c_{ki} x_k = f_i \qquad \text{pour} \quad i = 1, 2, \ldots, p$$

et

$$(3)\qquad \varphi(s) = f(s) + \mu \sum_{i=1}^{p} x_i a_i(s).$$

Nous pouvons remarquer que les solutions φ de l'équation (1) et les solutions x_i du système (4) se correspondent biunivoquement par les relations

$$\varphi(s) = f(s) + \mu \sum_{i=1}^{p} x_i a_i(s),$$

$$x_i = \int_a^b b_i(t)\varphi(t)\,dt.$$

De même, l'équation (1') est équivalente à

$$y_i - \mu \sum_{k=1}^{p} c_{ik} y_k = g_i \qquad \text{pour} \quad i = 1, 2, \ldots, p$$

et

$$\psi(s) = g(s) + \lambda \Sigma y_i b_i(s).$$

2. Discussion. — Nous devons introduire le déterminant du système (4) d'équations linéaires

$$D(\mu) = \begin{vmatrix} 1-\mu c_{11} & -\mu c_{21} & \ldots & -\mu c_{p1} \\ -\mu c_{12} & 1-\mu c_{22} & \ldots & -\mu c_{p2} \\ \ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots \\ -\mu c_{1p} & -\mu c_{2p} & \ldots & 1-\mu c_{pp} \end{vmatrix}.$$

Premier cas. — Si μ a une valeur telle que le déterminant $D(\mu)$ ne soit pas nul, le système (4) présentera une solution unique, ainsi que le système transposé. L'équation (1) de Fredholm présentera donc une solution φ et une seule donnée par (3); il en sera de même de l'équation transposée (1'). Les équations homogènes correspondantes (2) et (2') présenteront 0 pour une seule solution.

Deuxième cas. — Supposons que μ ait une valeur telle que le déterminant $D(\mu)$ soit nul. Considérons alors les systèmes homogènes

relatifs aux systèmes (4) et (4′)

$$(5)\qquad x_i - \mu \sum_{k=1}^{p} c_{ki} x_k = 0 \qquad \text{pour} \quad i = 1, 2, \ldots, p$$

et

$$(5')\qquad y_i - \mu \sum_{k=1}^{p} c_{ik} y_k = 0 \qquad \text{pour} \quad i = 1, 2, \ldots, p.$$

Ils ont le même nombre r de solutions linéairement indépendantes ([1]). Pour que le système (4) soit possible, il faut et il suffit que le vecteur f de composantes $(f_1, \ldots, f_p)$ soit orthogonal aux solutions de (5′), c'est-à-dire à r d'entre elles indépendantes

$$\sum_{i=1}^{p} f_i y_i^{(l)} = 0 \qquad \text{pour} \quad l = 1, 2, \ldots, r.$$

S'il en est ainsi, la solution générale de (4) dépend de r constantes arbitraires; elle est la somme d'une solution particulière de (4) et de la solution générale de (5).

Avant de conclure pour l'équation de Fredholm, vérifions que la *correspondance biunivoque entre les solutions* φ *de l'équation homogène* (2) *de Fredholm et les solutions* x_i *du système* (5) conserve la dépendance et l'indépendance linéaire. Supposons, en effet, que nous ayons

$$\sum_{k=1}^{j} c_k \varphi_k^{(s)} = 0.$$

Puisque, d'autre part

$$\varphi_k^{(s)} = \mu \sum_{i=1}^{p} x_i^{(k)} a_i(s),$$

([1]) Les solutions φ de l'équation homogène (2) et les solutions x_i du système (5) se correspondent biunivoquement par

$$\varphi = \mu \sum_{i=1}^{p} x_i a_i(s),$$

$$x_i = \int_a^b b_i \varphi \, dt.$$

Même remarque pour les solutions de (2′) au (5′).

la relation précédente s'écrit

$$\sum_{i=1}^{p} a_i(s) \sum_{k=1}^{j} c_k x_i^{(k)} = 0.$$

Et puisque les fonctions $a_i(s)$ sont linéairement indépendantes, ceci entraîne

$$\sum_{k=1}^{j} c_k x_i^{(k)} = 0 \qquad \text{pour} \quad i = 1, 2, \ldots, p,$$

c'est-à-dire

$$\sum_{k=1}^{j} c_k X^{(k)} = 0,$$

$X^{(k)}$ étant le vecteur de composantes $x_1^{(k)}, \ldots, x_p^{(k)}$.

Toute relation existant entre les solutions φ de (2) existe donc entre les solutions $X^{(k)}$ de (5). Et réciproquement, car ces calculs peuvent se refaire en sens inverse.

Par conséquent, si $D(\mu) = 0$, les équations (5) et (5′) présentent le même nombre r de solutions linéairement indépendantes ; les équations homogènes (2) et (2′) présentent ce même nombre r de solutions linéairement indépendantes.

L'équation (1) ne présentera de solution φ que si le système (4)

$$(4) \qquad x_i - \mu \sum_{k=1}^{p} c_{ki} x_k = f_i \qquad (i = 1, 2, \ldots, p)$$

en admet.

Si donc on désigne par $Y^{(1)}, \ldots, Y^{(r)}$ les r vecteurs solutions indépendantes du système transposé homogène

$$(5') \qquad y_i - \mu \sum_{k=1}^{p} c_{ik} y_k = 0,$$

les composantes de $Y^{(l)}$ étant $y_1^{(l)}, \ldots, y_p^{(l)}$ $(l = 1, 2, \ldots, r)$, la condition *nécessaire* et *suffisante* pour l'existence de solutions de (4) est que le vecteur $f\,(f_1, \ldots, f_p)$ soit orthogonal aux n vecteurs $Y^{(l)}$ c'est-à-dire que l'on ait

$$(6) \qquad \sum_{k=1}^{p} y_i^{(l)} f_i = 0 \qquad \text{pour} \quad l = 1, 2, \ldots, r.$$

Inversement, si (6) est réalisé, (4) admet une solution générale égale à une solution particulière augmentée de la solution générale de (5).

Or, à $Y^{(l)}$, correspond une solution $\psi^{(l)}(s)$ de (2'), équation de Fredholm homogène transposée, par

$$\psi^{(l)}(s) = \mu \sum_{i=1}^{p} y_i^{(l)} b_i(s),$$

les conditions (6) s'écrivent alors, à cause de $f_i = \int_a^b b_i(s) f(s)\, ds$,

$$0 = \sum_i y_i^{(l)} \int_a^b b_i f\, ds = \int_a^b f(s) \left(\sum_i y_i^{(l)} b_i \right) ds = \frac{1}{\mu} \int_a^b f(s) \psi^{(l)}(s)\, ds,$$

soit

$$(7) \qquad (f, \psi^{(l)}) = \int_a^b f(s) \psi^{(l)}(s)\, ds = 0 \qquad \text{pour} \quad l = 1, 2, \ldots, r.$$

Les conditions (7) expriment que $f(s)$ *est orthogonale à toutes les solutions de* (2'), équation homogène transposée.

Réciproquement, on voit que les conditions nécessaires et suffisantes (7) entraînent (6) auxquelles elles sont équivalentes.

Il faut donc et il suffit que le second membre f soit orthogonal aux r solutions linéairement indépendantes du système homogène transposé (2') S'il en est ainsi, la solution de (1) ne sera pas unique car la solution générale de (1) est la somme d'une solution particulière et de la solution générale de (2), en raison de la linéarité de l'équation de Fredholm, mise en évidence dans l'introduction.

3. **Conclusion. Théorème de Fredholm.** — Il peut s'énoncer sans l'aide des déterminants. Toutefois nous mettrons entre parenthèses ce qui est relatif aux déterminants, et le lecteur remarquera que le théorème conserve tout son sens lorsqu'on ne lit pas ce qui est entre parenthèses.

Un nombre μ étant donné, deux cas sont possibles :

a. Ou bien l'équation de Fredholm (1) admet, quel que soit le second membre f, une solution et une seule φ; en particulier l'équation homogène n'admet que la solution 0. Dans ce cas $[D(\mu) \neq 0]$ les mêmes résultats s'appliquent à l'équation transposée et à l'équation homogène transposée.

b. Ou bien l'équation homogène admet un nombre $r \neq 0$ de solutions linéairement indépendantes $\varphi^{(i)}$, l'équation homogène transposée admet

le même nombre r de solutions linéairement indépendantes $\psi^{(i)}$. La condition nécessaire et suffisante pour que l'équation (1) de Fredholm soit résoluble est que l'on ait $(f, \psi^{(i)}) = 0$ pour $i = 1, 2, \ldots, r$; et si elle est résoluble sa solution générale φ est égale à la somme d'une solution particulière φ^0 et de la solution générale $\sum_{i=1}^{r} c_i \varphi^{(i)}$ de l'équation homogène. [Dans ce cas $D(\mu) = 0$]

$$\varphi = \varphi^0 + \sum_{i=1}^{r} c_i \varphi^{(i)} \quad (^1),$$

Les valeurs μ qui annulent $D(\mu)$, polynome en μ de degré p, sont les *valeurs caractéristiques* (*ou propres*) *de l'équation* (1) ou (1′) ou (2) ou (2′).

4. **Remarque.** — Les noyaux de Goursat se sont révélés utiles même pour l'étude de *l'équation générale de deuxième espèce à noyau* $k(s, t)$ *continu* (et même plus généraux). Il suffit pour cela de remarquer que tout noyau continu $k(s, t)$ peut, suivant un théorème de Weierstrass, se développer *en série uniformément convergente de polynomes en* (s, t) ou être *arbitrairement approché* dans le domaine $a \leqq s \leqq b$, $a \leqq t \leqq b$, par un polynome en (s, t). En substituant à $k(s, t)$ dans (1) un polynome en (s, t) approché, on a une équation approchée de (1) dont les solutions approchent celles de (1). Or un polynome en (s, t) est un noyau de Goursat, nous savons donc discuter la résolution de l'équation approchée et trouver les solutions par ce qui précède. Il est possible, en prenant des approximations de $k(s, t)$ de plus en plus serrées, de démontrer l'existence des solutions de (1) et d'approcher de plus en plus ces solutions. Nous nous bornerons à ces indications, renvoyant sur ce sujet aux Mémoires de M. Goursat (*loc. cit*) et de M. Lebesgue (*Bull. Soc. Math. de Fr.*, 1908); *voir* aussi Courant et Hilbert, *Méthoden der mathematischen Physik*, Chap. 3].

---

(1) Dans cette solution générale, on peut choisir une solution unique particulièrement intéressante en exigeant de φ la relation $(\varphi, \varphi^{(i)}) = 0$. Les équations en c_i ont alors un déterminant qui est un déterminant de Gram, lequel est $\neq 0$, puisque les $\varphi^{(i)}$ sont indépendants.

FIN.

TABLE DES MATIÈRES.

APERÇU HISTORIQUE.

CHAPITRE I.

CHAPITRE II.

CHAPITRE III.

CHAPITRE IV.

Étude particulière des matrices et opérateurs hermitiens.

CHAPITRE V.

Valeurs d'une forme associée à un opérateur. Lien avec la réduction canonique d'une forme hermitienne (Méthode d'extremum pour la réduction). Domaine des valeurs d'une forme associée à un opérateur quelconque.

CHAPITRE VI.

102825 Paris. — Imprimerie GAUTHIER-VILLARS, quai des Grands-Augustins, 55